(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 September 2001 (27.09.2001)

PCT

(10) International Publication Number
WO 01/71004 A2

(51) International Patent Classification[7]: C12N 15/57, 9/64, 9/48, 1/21, 5/10, A01K 67/027, C12N 15/00, C07K 16/40, C12Q 1/68, A61K 38/48, C12Q 1/37

(21) International Application Number: PCT/US01/08441

(22) International Filing Date: 16 March 2001 (16.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/190,708	17 March 2000 (17.03.2000)	US
60/193,182	30 March 2000 (30.03.2000)	US
60/197,086	14 April 2000 (14.04.2000)	US
60/199,022	20 April 2000 (20.04.2000)	US
60/200,227	28 April 2000 (28.04 2000)	US

(71) Applicant *(for all designated States except US)*: **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **POLICKY, Jennifer, L.** [US/US]; 4864 National Avenue, San Jose, CA 95124 (US). **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **KHAN, Farrah, A.** [IN/US]; 333 Escuela Avenue #221, Mountain View, CA 94040 (US). **AU-YOUNG, Janice** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **HILLMAN, Jennifer, L.** [US/US]; 230 Monroe Drive #17, Mountain View, CA 94040 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street, #25, San Leandro, CA 94577 (US). **HE, Ann** [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US). **HAFALIA, April** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/71004 A2

(54) Title: PROTEASES

(57) Abstract: The invention provides human proteases (PRTS) and polynucleotides which identify and encode PRTS. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PRTS.

PROTEASES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of proteases and to the use of these sequences in the diagnosis, treatment, and prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

BACKGROUND OF THE INVENTION

Proteases cleave proteins and peptides at the peptide bond that forms the backbone of the protein or peptide chain. Proteolysis is one of the most important and frequent enzymatic reactions that occurs both within and outside of cells. Proteolysis is responsible for the activation and maturation of nascent polypeptides, the degradation of misfolded and damaged proteins, and the controlled turnover of peptides within the cell. Proteases participate in digestion, endocrine function, and tissue remodeling during embryonic development, wound healing, and normal growth. Proteases can play a role in regulatory processes by affecting the half life of regulatory proteins. Proteases are involved in the etiology or progression of disease states such as inflammation, angiogenesis, tumor dispersion and metastasis, cardiovascular disease, neurological disease, and bacterial, parasitic, and viral infections.

Proteases can be categorized on the basis of where they cleave their substrates. Exopeptidases, which include aminopeptidases, dipeptidyl peptidases, tripeptidases, carboxypeptidases, peptidyl-di-peptidases, dipeptidases, and omega peptidases, cleave residues at the termini of their substrates. Endopeptidases, including serine proteases, cysteine proteases, and metalloproteases, cleave at residues within the peptide. Four principal categories of mammalian proteases have been identified based on active site structure, mechanism of action, and overall three-dimensional structure. (See Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp. 1-5.)

Serine Proteases

The serine proteases (SPs) are a large, widespread family of proteolytic enzymes that include the digestive enzymes trypsin and chymotrypsin, components of the complement and blood-clotting cascades, and enzymes that control the degradation and turnover of macromolecules within the cell and in the extracellular matrix. Most of the more than 20 subfamilies can be grouped into six clans, each with a common ancestor. These six clans are hypothesized to have descended from at least four evolutionarily distinct ancestors. SPs are named for the presence of a serine residue found in the

active catalytic site of most families. The active site is defined by the catalytic triad, a set of conserved asparagine, histidine, and serine residues critical for catalysis. These residues form a charge relay network that facilitates substrate binding. Other residues outside the active site form an oxyanion hole that stabilizes the tetrahedral transition intermediate formed during catalysis. SPs have a wide range of substrates and can be subdivided into subfamilies on the basis of their substrate specificity. The main subfamilies are named for the residue(s) after which they cleave: trypases (after arginine or lysine), aspases (after aspartate), chymases (after phenylalanine or leucine), metases (methionine), and serases (after serine) (Rawlings, N.D. and A.J. Barrett (1994) Meth. Enzymol. 244:19-61).

Most mammalian serine proteases are synthesized as zymogens, inactive precursors that are activated by proteolysis. For example, trypsinogen is converted to its active form, trypsin, by enteropeptidase. Enteropeptidase is an intestinal protease that removes an N-terminal fragment from trypsinogen. The remaining active fragment is trypsin, which in turn activates the precursors of the other pancreatic enzymes. Likewise, proteolysis of prothrombin, the precursor of thrombin, generates three separate polypeptide fragments. The N-terminal fragment is released while the other two fragments, which comprise active thrombin, remain associated through disulfide bonds.

The two largest SP subfamilies are the chymotrypsin (S1) and subtilisin (S8) families. Some members of the chymotrypsin family contain two structural domains unique to this family. Kringle domains are triple-looped, disulfide cross-linked domains found in varying copy number. Kringles are thought to play a role in binding mediators such as membranes, other proteins or phospholipids, and in the regulation of proteolytic activity (PROSITE PDOC00020). Apple domains are 90 amino-acid repeated domains, each containing six conserved cysteines. Three disulfide bonds link the first and sixth, second and fifth, and third and fourth cysteines (PROSITE PDOC00376). Apple domains are involved in protein-protein interactions. S1 family members include trypsin, chymotrypsin, coagulation factors IX-XII, complement factors B, C, and D, granzymes, kallikrein, and tissue- and urokinase-plasminogen activators. The subtilisin family has members found in the eubacteria, archaebacteria, eukaryotes, and viruses. Subtilisins include the proprotein-processing endopeptidases kexin and furin and the pituitary prohormone convertases PC1, PC2, PC3, PC6, and PACE4 (Rawlings and Barrett, supra).

SPs have functions in many normal processes and some have been implicated in the etiology or treatment of disease. Enterokinase, the initiator of intestinal digestion, is found in the intestinal brush border, where it cleaves the acidic propeptide from trypsinogen to yield active trypsin (Kitamoto, Y. et al. (1994) Proc. Natl. Acad. Sci. USA 91:7588-7592). Prolylcarboxypeptidase, a lysosomal serine peptidase that cleaves peptides such as angiotensin II and III and [des-Arg9] bradykinin, shares sequence homology with members of both the serine carboxypeptidase and

prolylendopeptidase families (Tan, F. et al. (1993) J. Biol. Chem. 268:16631-16638). The protease neuropsin may influence synapse formation and neuronal connectivity in the hippocampus in response to neural signaling (Chen, Z.-L. et al. (1995) J Neurosci 15:5088-5097). Tissue plasminogen activator is useful for acute management of stroke (Zivin, J.A. (1999) Neurology 53:14-19) and myocardial infarction (Ross, A.M. (1999) Clin. Cardiol. 22:165-171). Some receptors (PAR, for proteinase-activated receptor), highly expressed throughout the digestive tract, are activated by proteolytic cleavage of an extracellular domain. The major agonists for PARs, thrombin, trypsin, and mast cell tryptase, are released in allergy and inflammatory conditions. Control of PAR activation by proteases has been suggested as a promising therapeutic target (Vergnolle, N. (2000) Aliment. Pharmacol. Ther. 14:257-266; Rice, K.D. et al. (1998) Curr. Pharm. Des. 4:381-396). Prostate-specific antigen (PSA) is a kallikrein-like serine protease synthesized and secreted exclusively by epithelial cells in the prostate gland. Serum PSA is elevated in prostate cancer and is the most sensitive physiological marker for monitoring cancer progression and response to therapy. PSA can also identify the prostate as the origin of a metastatic tumor (Brawer, M.K. and P.H. Lange (1989) Urology 33:11-16).

The signal peptidase is a specialized class of SP found in all prokaryotic and eukaryotic cell types that serves in the processing of signal peptides from certain proteins. Signal peptides are amino-terminal domains of a protein which direct the protein from its ribosomal assembly site to a particular cellular or extracellular location. Once the protein has been exported, removal of the signal sequence by a signal peptidase and posttranslational processing, e.g., glycosylation or phosphorylation, activate the protein. Signal peptidases exist as multi-subunit complexes in both yeast and mammals. The canine signal peptidase complex is composed of five subunits, all associated with the microsomal membrane and containing hydrophobic regions that span the membrane one or more times (Shelness, G.S. and G. Blobel (1990) J. Biol. Chem. 265:9512-9519). Some of these subunits serve to fix the complex in its proper position on the membrane while others contain the actual catalytic activity.

Another family of proteases which have a serine in their active site are dependent on the hydrolysis of ATP for their activity. These proteases contain proteolytic core domains and regulatory ATPase domains which can be identified by the presence of the P-loop, an ATP/GTP-binding motif (PROSITE PDOC00803). Members of this family include the eukaryotic mitochondrial matrix proteases, Clp protease and the proteasome. Clp protease was originally found in plant chloroplasts but is believed to be widespread in both prokaryotic and eukaryotic cells. The gene for early-onset torsion dystonia encodes a protein related to Clp protease (Ozelius, L.J. et al. (1998) Adv. Neurol. 78:93-105).

The proteasome is an intracellular protease complex found in some bacteria and in all

eukaryotic cells, and plays an important role in cellular physiology. Proteasomes are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins of all types, including proteins that function to activate or repress cellular processes such as transcription and cell cycle progression (Ciechanover, A. (1994) Cell 79:13-21). In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin, a small heat stable protein. The ubiquitinated protein is then recognized and degraded by the proteasome. The resultant ubiquitin-peptide complex is hydrolyzed by a ubiquitin carboxyl terminal hydrolase, and free ubiquitin is released for reutilization by the UCS. Ubiquitin-proteasome systems are implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes (p53), cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, supra). This pathway has been implicated in a number of diseases, including cystic fibrosis, Angelman's syndrome, and Liddle syndrome (reviewed in Schwartz, A.L. and A. Ciechanover (1999) Annu. Rev. Med. 50:57-74). A murine proto-oncogene, Unp, encodes a nuclear ubiquitin protease whose overexpression leads to oncogenic transformation of NIH3T3 cells. The human homologue of this gene is consistently elevated in small cell tumors and adenocarcinomas of the lung (Gray, D.A. (1995) Oncogene 10:2179-2183). Ubiquitin carboxyl terminal hydrolase is involved in the differentiation of a lymphoblastic leukemia cell line to a non-dividing mature state (Maki, A. et al. (1996) Differentiation 60:59-66). In neurons, ubiquitin carboxyl terminal hydrolase (PGP 9.5) expression is strong in the abnormal structures that occur in human neurodegenerative diseases (Lowe, J. et al. (1990) J. Pathol. 161:153-160). The proteasome is a large (~2000 kDa) multisubunit complex composed of a central catalytic core containing a variety of proteases arranged in four seven-membered rings with the active sites facing inwards into the central cavity, and terminal ATPase subunits covering the outer port of the cavity and regulating substrate entry (for review, see Schmidt, M. et al. (1999) Curr. Opin. Chem. Biol. 3:584-591).

Cysteine Proteases

Cysteine proteases (CPs) are involved in diverse cellular processes ranging from the processing of precursor proteins to intracellular degradation. Nearly half of the CPs known are present only in viruses. CPs have a cysteine as the major catalytic residue at the active site where catalysis proceeds via a thioester intermediate and is facilitated by nearby histidine and asparagine residues. A glutamine residue is also important, as it helps to form an oxyanion hole. Two important CP families include the papain-like enzymes (C1) and the calpains (C2). Papain-like family members are generally lysosomal or secreted and therefore are synthesized with signal peptides as well as propeptides. Most members bear a conserved motif in the propeptide that may have structural significance (Karrer, K.M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3063-3067). Three-dimensional structures of papain family members show a bilobed molecule with the catalytic site

located between the two lobes. Papains include cathepsins B, C, H, L, and S, certain plant allergens and dipeptidyl peptidase (for a review, see Rawlings, N.D. and A.J. Barrett (1994) Meth. Enzymol. 244:461-486).

Some CPs are expressed ubiquitously, while others are produced only by cells of the immune system. Of particular note, CPs are produced by monocytes, macrophages and other cells which migrate to sites of inflammation and secrete molecules involved in tissue repair. Overabundance of these repair molecules plays a role in certain disorders. In autoimmune diseases such as rheumatoid arthritis, secretion of the cysteine peptidase cathepsin C degrades collagen, laminin, elastin and other structural proteins found in the extracellular matrix of bones. Bone weakened by such degradation is also more susceptible to tumor invasion and metastasis. Cathepsin L expression may also contribute to the influx of mononuclear cells which exacerbates the destruction of the rheumatoid synovium (Keyszer, G.M. (1995) Arthritis Rheum. 38:976-984).

Calpains are calcium-dependent cytosolic endopeptidases which contain both an N-terminal catalytic domain and a C-terminal calcium-binding domain. Calpain is expressed as a proenzyme heterodimer consisting of a catalytic subunit unique to each isoform and a regulatory subunit common to different isoforms. Each subunit bears a calcium-binding EF-hand domain. The regulatory subunit also contains a hydrophobic glycine-rich domain that allows the enzyme to associate with cell membranes. Calpains are activated by increased intracellular calcium concentration, which induces a change in conformation and limited autolysis. The resultant active molecule requires a lower calcium concentration for its activity (Chan, S.L. and M.P. Mattson (1999) J. Neurosci. Res. 58:167-190). Calpain expression is predominantly neuronal, although it is present in other tissues. Several chronic neurodegenerative disorders, including ALS, Parkinson's disease and Alzheimer's disease are associated with increased calpain expression (Chan and Mattson, supra). Calpain-mediated breakdown of the cytoskeleton has been proposed to contribute to brain damage resulting from head injury (McCracken, E. et al. (1999) J. Neurotrauma 16:749-761). Calpain-3 is predominantly expressed in skeletal muscle, and is responsible for limb-girdle muscular dystrophy type 2A (Minami, N. et al. (1999) J. Neurol. Sci. 171:31-37).

Another family of thiol proteases is the caspases, which are involved in the initiation and execution phases of apoptosis. A pro-apoptotic signal can activate initiator caspases that trigger a proteolytic caspase cascade, leading to the hydrolysis of target proteins and the classic apoptotic death of the cell. Two active site residues, a cysteine and a histidine, have been implicated in the catalytic mechanism. Caspases are among the most specific endopeptidases, cleaving after aspartate residues. Caspases are synthesized as inactive zymogens consisting of one large (p20) and one small (p10) subunit separated by a small spacer region, and a variable N-terminal prodomain. This prodomain interacts with cofactors that can positively or negatively affect apoptosis. An activating

signal causes autoproteolytic cleavage of a specific aspartate residue (D297 in the caspase-1 numbering convention) and removal of the spacer and prodomain, leaving a p10/p20 heterodimer. Two of these heterodimers interact via their small subunits to form the catalytically active tetramer. The long prodomains of some caspase family members have been shown to promote dimerization and auto-processing of procaspases. Some caspases contain a "death effector domain" in their prodomain by which they can be recruited into self-activating complexes with other caspases and FADD protein associated death receptors or the TNF receptor complex. In addition, two dimers from different caspase family members can associate, changing the substrate specificity of the resultant tetramer. Endogenous caspase inhibitors (inhibitor of apoptosis proteins, or IAPs) also exist. All these interactions have clear effects on the control of apoptosis (reviewed in Chan and Mattson, supra; Salveson, G.S. and V.M. Dixit (1999) Proc. Natl. Acad. Sci. USA 96:10964-10967).

Caspases have been implicated in a number of diseases. Mice lacking some caspases have severe nervous system defects due to failed apoptosis in the neuroepithelium and suffer early lethality. Others show severe defects in the inflammatory response, as caspases are responsible for processing IL-1b and possibly other inflammatory cytokines (Chan and Mattson, supra). Cowpox virus and baculoviruses target caspases to avoid the death of their host cell and promote successful infection. In addition, increases in inappropriate apoptosis have been reported in AIDS, neurodegenerative diseases and ischemic injury, while a decrease in cell death is associated with cancer (Salveson and Dixit, supra; Thompson, C.B. (1995) Science 267:1456-1462).

Aspartyl proteases

Aspartyl proteases (APs) include the lysosomal proteases cathepsins D and E, as well as chymosin, renin, and the gastric pepsins. Most retroviruses encode an AP, usually as part of the pol polyprotein. APs, also called acid proteases, are monomeric enzymes consisting of two domains, each domain containing one half of the active site with its own catalytic aspartic acid residue. APs are most active in the range of pH 2–3, at which one of the aspartate residues is ionized and the other neutral. The pepsin family of APs contains many secreted enzymes, and all are likely to be synthesized with signal peptides and propeptides. Most family members have three disulfide loops, the first ~5 residue loop following the first aspartate, the second 5-6 residue loop preceding the second aspartate, and the third and largest loop occurring toward the C terminus. Retropepsins, on the other hand, are analogous to a single domain of pepsin, and become active as homodimers with each retropepsin monomer contributing one half of the active site. Retropepsins are required for processing the viral polyproteins.

APs have roles in various tissues, and some have been associated with disease. Renin mediates the first step in processing the hormone angiotensin, which is responsible for regulating electrolyte balance and blood pressure (reviewed in Crews, D.E. and S.R. Williams (1999) Hum.

Biol. 71:475-503). Abnormal regulation and expression of cathepsins are evident in various inflammatory disease states. Expression of cathepsin D is elevated in synovial tissues from patients with rheumatoid arthritis and osteoarthritis. The increased expression and differential regulation of the cathepsins are linked to the metastatic potential of a variety of cancers (Chambers, A.F. et al. (1993) Crit. Rev. Oncol. 4:95-114).

Metalloproteases

Metalloproteases require a metal ion for activity, usually manganese or zinc. Examples of manganese metalloenzymes include aminopeptidase P and human proline dipeptidase (PEPD). Aminopeptidase P can degrade bradykinin, a nonapeptide activated in a variety of inflammatory responses. Aminopeptidase P has been implicated in coronary ischemia/reperfusion injury. Administration of aminopeptidase P inhibitors has been shown to have a cardioprotective effect in rats (Ersahin, C. et al (1999) J. Cardiovasc. Pharmacol. 34:604-611).

Most zinc-dependent metalloproteases share a common sequence in the zinc-binding domain. The active site is made up of two histidines which act as zinc ligands and a catalytic glutamic acid C-terminal to the first histidine. Proteins containing this signature sequence are known as the metzincins and include aminopeptidase N, angiotensin-converting enzyme, neurolysin, the matrix metalloproteases and the adamalysins (ADAMS). An alternate sequence is found in the zinc carboxypeptidases, in which all three conserved residues – two histidines and a glutamic acid – are involved in zinc binding.

A number of the neutral metalloendopeptidases, including angiotensin converting enzyme and the aminopeptidases, are involved in the metabolism of peptide hormones. High aminopeptidase B activity, for example, is found in the adrenal glands and neurohypophyses of hypertensive rats (Prieto, I. et al. (1998) Horm. Metab. Res. 30:246-248). Oligopeptidase M/neurolysin can hydrolyze bradykinin as well as neurotensin (Serizawa, A. et al. (1995) J. Biol. Chem 270:2092-2098). Neurotensin is a vasoactive peptide that can act as a neurotransmitter in the brain, where it has been implicated in limiting food intake (Tritos, N.A. et al. (1999) Neuropeptides 33:339-349).

The matrix metalloproteases (MMPs) are a family of at least 23 enzymes that can degrade components of the extracellular matrix (ECM). They are Zn^{+2} endopeptidases with an N-terminal catalytic domain. Nearly all members of the family have a hinge peptide and C-terminal domain which can bind to substrate molecules in the ECM or to inhibitors produced by the tissue (TIMPs, for tissue inhibitor of metalloprotease; Campbell, I.L. et al. (1999) Trends Neurosci. 22:285). The presence of fibronectin-like repeats, transmembrane domains, or C-terminal hemopexinase-like domains can be used to separate MMPs into collagenase, gelatinase, stromelysin and membrane-type MMP subfamilies. In the inactive form, the Zn^{+2} ion in the active site interacts with a cysteine in the pro-sequence. Activating factors disrupt the Zn^{+2}-cysteine interaction, or "cysteine switch," exposing

the active site. This partially activates the enzyme, which then cleaves off its propeptide and becomes fully active. MMPs are often activated by the serine proteases plasmin and furin. MMPs are often regulated by stoichiometric, noncovalent interactions with inhibitors; the balance of protease to inhibitor, then, is very important in tissue homeostasis (reviewed in Yong, V.W. et al. (1998) Trends Neurosci. 21:75).

MMPs are implicated in a number of diseases including osteoarthritis (Mitchell, P. et al. (1996) J. Clin. Invest. 97:761), atherosclerotic plaque rupture (Sukhova, G.K. et al. (1999) Circulation 99:2503), aortic aneurysm (Schneiderman, J. et al. (1998) Am. J. Path. 152:703), non-healing wounds (Saarialho-Kere, U.K. et al. (1994) J. Clin. Invest. 94:79), bone resorption (Blavier, L. and J.M. Delaisse (1995) J. Cell Sci. 108:3649), age-related macular degeneration (Steen, B. et al. (1998) Invest. Ophthalmol. Vis. Sci. 39:2194), emphysema (Finlay, G.A. et al. (1997) Thorax 52:502), myocardial infarction (Rohde, L.E. et al. (1999) Circulation 99:3063) and dilated cardiomyopathy (Thomas, C.V. et al. (1998) Circulation 97:1708). MMP inhibitors prevent metastasis of mammary carcinoma and experimental tumors in rat, and Lewis lung carcinoma, hemangioma, and human ovarian carcinoma xenografts in mice (Eccles, S.A. et al. (1996) Cancer Res. 56:2815; Anderson et al. (1996) Cancer Res. 56:715-718; Volpert, O.V. et al. (1996) J. Clin. Invest. 98:671; Taraboletti, G. et al. (1995) J. NCI 87:293; Davies, B. et al. (1993) Cancer Res. 53:2087). MMPs may be active in Alzheimer's disease. A number of MMPs are implicated in multiple sclerosis, and administration of MMP inhibitors can relieve some of its symptoms (reviewed in Yong, supra).

Another family of metalloproteases is the ADAMs, for A Disintegrin and Metalloprotease Domain, which they share with their close relatives the adamalysins, snake venom metalloproteases (SVMPs). ADAMs combine features of both cell surface adhesion molecules and proteases, containing a prodomain, a protease domain, a disintegrin domain, a cysteine rich domain, an epidermal growth factor repeat, a transmembrane domain, and a cytoplasmic tail. The first three domains listed above are also found in the SVMPs. The ADAMs possess four potential functions: proteolysis, adhesion, signaling and fusion. The ADAMs share the metzincin zinc binding sequence and are inhibited by some MMP antagonists such as TIMP-1.

ADAMs are implicated in such processes as sperm-egg binding and fusion, myoblast fusion, and protein-ectodomain processing or shedding of cytokines, cytokine receptors, adhesion proteins and other extracellular protein domains (Schlöndorff, J. and C.P. Blobel (1999) J. Cell. Sci. 112:3603-3617). The Kuzbanian protein cleaves a substrate in the NOTCH pathway (possibly NOTCH itself), activating the program for lateral inhibition in Drosophila neural development. Two ADAMs, TACE (ADAM 17) and ADAM 10, are proposed to have analogous roles in the processing of amyloid precursor protein in the brain (Schlöndorff and Blobel, supra). TACE has also been

identified as the TNF activating enzyme (Black, R.A. et al. (1997) Nature 385:729). TNF is a pleiotropic cytokine that is important in mobilizing host defenses in response to infection or trauma, but can cause severe damage in excess and is often overproduced in autoimmune disease. TACE cleaves membrane-bound pro-TNF to release a soluble form. Other ADAMs may be involved in a similar type of processing of other membrane-bound molecules.

The ADAMTS sub-family has all of the features of ADAM family metalloproteases and contain an additional thrombospondin domain (TS). The prototypic ADAMTS was identified in mouse, found to be expressed in heart and kidney and upregulated by proinflammatory stimuli (Kuno, K. et al. (1997) J. Biol. Chem. 272:556). To date eleven members are recognized by the Human Genome Organization (HUGO; http://www.gene.ucl.ac.uk/users/hester/adamts.html#Approved). Members of this family have the ability to degrade aggrecan, a high molecular weight proteoglycan which provides cartilage with important mechanical properties including compressibility, and which is lost during the development of arthritis. Enzymes which degrade aggrecan are thus considered attractive targets to prevent and slow the degradation of articular cartilage (See, e.g., Tortorella, M.D. (1999) Science 284:1664; Abbaszade, I. (1999) J. Biol. Chem. 274:23443). Other members are reported to have antiangiogenic potential (Kuno et al., supra) and/or procollagen processing (Colige, A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2374).

The discovery of new proteases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of proteases.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, proteases, referred to collectively as "PRTS" and individually as "PRTS-1," "PRTS-2," "PRTS-3," "PRTS-4," "PRTS-5," "PRTS-6," "PRTS-7," "PRTS-8," "PRTS-9," "PRTS-10," and "PRTS-11." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-11. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:12-22.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to

a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the

invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to a polypeptide sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:12-22, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"PRTS" refers to the amino acid sequences of substantially purified PRTS obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of PRTS. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

An "allelic variant" is an alternative form of the gene encoding PRTS. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or

many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PRTS include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PRTS or a polypeptide with at least one functional characteristic of PRTS. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PRTS, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding PRTS. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PRTS. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PRTS is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of PRTS. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PRTS either by directly interacting with PRTS or by acting on components of the biological pathway in which PRTS participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant.

Antibodies that bind PRTS polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PRTS, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PRTS or fragments of PRTS may be

employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of PRTS or the polynucleotide encoding PRTS which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:12-22 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:12-22, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:12-22 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:12-22 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:12-22 and the region of SEQ ID NO:12-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-11 is encoded by a fragment of SEQ ID NO:12-22. A fragment of SEQ ID NO:1-11 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-11. For example, a fragment of SEQ ID NO:1-11 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-11. The precise length of a fragment of SEQ ID NO:1-11 and the region of SEQ ID NO:1-11 to which the fragment

corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of

the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PRTS which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PRTS which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of PRTS. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PRTS.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of a PRTS may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PRTS.

"Probe" refers to nucleic acid sequences encoding PRTS, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers

may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the

artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing PRTS, nucleic acids encoding PRTS, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free,

preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-

1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human proteases (PRTS), the polynucleotides encoding PRTS, and the use of these compositions for the diagnosis, treatment, or prevention of gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are proteases. For example, SEQ ID NO:1 is 73% identical, from residue M1 to residue L717, to human precursor of P100 serine protease of Ra-reactive factor (GenBank ID g439713) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.50E-271, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a trypsin domain and two CUB domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is a serine protease (note that "serine protease" encompasses a widespread family of proteolytic enzymes including the digestive enzyme trypsin). SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-11 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence

identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:12-22 or that distinguish between SEQ ID NO:12-22 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5′) and stop (3′) positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 3741323H1 is the identification number of an Incyte cDNA sequence, and MENTNOT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 70885185V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g2821051) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6693583_012, is the identification number of a Genscan-predicted coding sequence, with g6693583 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, FL333966_00001 represents a "stitched" sequence in which 333966 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which

were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PRTS variants. A preferred PRTS variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PRTS amino acid sequence, and which contains at least one functional or structural characteristic of PRTS.

The invention also encompasses polynucleotides which encode PRTS. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22, which encodes PRTS. The polynucleotide sequences of SEQ ID NO:12-22, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding PRTS. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PRTS. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:12-22. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PRTS.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PRTS, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PRTS, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PRTS and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PRTS under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PRTS or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which

particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding PRTS and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PRTS and PRTS derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PRTS or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:12-22 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PRTS may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown

sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode PRTS may be cloned in recombinant DNA molecules that direct expression of PRTS, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PRTS.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PRTS-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PRTS, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding PRTS may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PRTS itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PRTS, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PRTS, the nucleotide sequences encoding PRTS or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PRTS. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PRTS. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PRTS and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PRTS and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PRTS. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO

J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PRTS. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PRTS can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding PRTS into the vector's multiple cloning site disrupts the *lac*Z gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PRTS are needed, e.g. for the production of antibodies, vectors which direct high level expression of PRTS may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PRTS. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PRTS. Transcription of sequences encoding PRTS may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or

pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PRTS may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PRTS in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PRTS in cell lines is preferred. For example, sequences encoding PRTS can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase. respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins

(GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PRTS is inserted within a marker gene sequence, transformed cells containing sequences encoding PRTS can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PRTS under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PRTS and that express PRTS may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PRTS using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PRTS is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding PRTS include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PRTS, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega

(Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding PRTS may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PRTS may be designed to contain signal sequences which direct secretion of PRTS through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PRTS may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PRTS protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PRTS activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PRTS encoding sequence and the heterologous protein sequence, so that PRTS may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10).

A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PRTS may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

PRTS of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PRTS. At least one and up to a plurality of test compounds may be screened for specific binding to PRTS. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PRTS, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PRTS binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PRTS, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing PRTS or cell membrane fractions which contain PRTS are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PRTS or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with PRTS, either in solution or affixed to a solid support, and detecting the binding of PRTS to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PRTS of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PRTS. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PRTS activity, wherein PRTS is combined with at least one test compound, and the activity of PRTS in the

presence of a test compound is compared with the activity of PRTS in the absence of the test compound. A change in the activity of PRTS in the presence of the test compound is indicative of a compound that modulates the activity of PRTS. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PRTS under conditions suitable for PRTS activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PRTS may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PRTS or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PRTS may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PRTS can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PRTS is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease.

Alternatively, a mammal inbred to overexpress PRTS, e.g., by secreting PRTS in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of PRTS and proteases. In addition, the expression of PRTS is closely associated with reproductive tissue, prostate cancer tissue, diseased, cancerous, and tumorous tissues, cardiovascular tissue, promonocytes, and with esophageal and seminal vesicle tissues. Therefore, PRTS appears to play a role in gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders. In the treatment of disorders associated with increased PRTS expression or activity, it is desirable to decrease the expression or activity of PRTS. In the treatment of disorders associated with decreased PRTS expression or activity, it is desirable to increase the expression or activity of PRTS.

Therefore, in one embodiment, PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular

calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma,

acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation

syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia.

In another embodiment, a vector capable of expressing PRTS or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified PRTS in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of PRTS may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PRTS including, but not limited to, those listed above.

In a further embodiment, an antagonist of PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS. Examples of such disorders include, but are not limited to, those gastrointestinal, cardiovascular, autoimmune/inflammatory, cell proliferative, developmental, epithelial, neurological, and reproductive disorders described above. In one aspect, an antibody which specifically binds PRTS may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PRTS.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PRTS may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PRTS including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PRTS may be produced using methods which are generally known in the art. In particular, purified PRTS may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PRTS. Antibodies to PRTS may also

be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PRTS or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PRTS have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PRTS amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PRTS may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce PRTS-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as

disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for PRTS may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PRTS and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PRTS epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PRTS. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of PRTS-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PRTS epitopes, represents the average affinity, or avidity, of the antibodies for PRTS. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular PRTS epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the PRTS-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PRTS, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PRTS-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and

guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PRTS, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PRTS. Such technology is well-known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PRTS. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) J. Allergy Cli. Immunol. 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PRTS may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399),

hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PRTS expression or regulation causes disease, the expression of PRTS from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PRTS are treated by constructing mammalian expression vectors encoding PRTS and introducing these vectors by mechanical means into PRTS-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PRTS include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PRTS may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PRTS from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PRTS expression are treated by constructing a retrovirus vector consisting of (i) the

polynucleotide encoding PRTS under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PRTS to cells which have one or more genetic abnormalities with respect to the expression of PRTS. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PRTS to target cells which have one or more genetic abnormalities with respect to the expression of PRTS. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PRTS to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has

been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PRTS to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PRTS into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PRTS-coding RNAs and the synthesis of high levels of PRTS in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PRTS into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful

because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PRTS.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PRTS. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PRTS. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PRTS expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PRTS may be therapeutically useful, and in the treament of disorders associated with decreased PRTS expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PRTS may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PRTS is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PRTS are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PRTS. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide

nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of PRTS, antibodies to PRTS, and mimetics, agonists, antagonists, or inhibitors of PRTS.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising PRTS or fragments thereof. For example, liposome preparations

containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PRTS or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PRTS or fragments thereof, antibodies of PRTS, and agonists, antagonists or inhibitors of PRTS, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind PRTS may be used for the diagnosis of disorders characterized by expression of PRTS, or in assays to monitor patients being treated with PRTS or agonists, antagonists, or inhibitors of PRTS. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PRTS include methods which utilize the antibody and a label to detect PRTS in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PRTS, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of PRTS expression. Normal or standard values for PRTS expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PRTS under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PRTS expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PRTS may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PRTS may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PRTS, and to monitor regulation of PRTS levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PRTS or closely related molecules may be used to identify nucleic acid sequences which encode PRTS. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PRTS, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PRTS encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:12-22 or from genomic sequences including promoters, enhancers, and introns of the PRTS gene.

Means for producing specific hybridization probes for DNAs encoding PRTS include the cloning of polynucleotide sequences encoding PRTS or PRTS derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PRTS may be used for the diagnosis of disorders associated with expression of PRTS. Examples of such disorders include, but are not limited to, a gastrointestinal disorder, such as dysphagia, peptic esophagitis, esophageal spasm, esophageal stricture, esophageal carcinoma, dyspepsia, indigestion, gastritis, gastric carcinoma, anorexia, nausea, emesis, gastroparesis, antral or pyloric edema, abdominal angina, pyrosis, gastroenteritis, intestinal obstruction, infections of the intestinal tract, peptic ulcer, cholelithiasis, cholecystitis, cholestasis, pancreatitis, pancreatic carcinoma, biliary tract disease, hepatitis, hyperbilirubinemia, cirrhosis, passive congestion of the liver, hepatoma, infectious colitis, ulcerative colitis, ulcerative proctitis, Crohn's disease, Whipple's disease, Mallory-Weiss syndrome, colonic carcinoma, colonic obstruction, irritable bowel syndrome, short bowel syndrome, diarrhea, constipation, gastrointestinal hemorrhage, acquired immunodeficiency syndrome (AIDS) enteropathy, jaundice, hepatic encephalopathy, hepatorenal syndrome, hepatic steatosis, hemochromatosis, Wilson's disease, alpha$_1$-antitrypsin deficiency, Reye's syndrome, primary sclerosing cholangitis, liver infarction, portal vein obstruction and thrombosis, centrilobular necrosis, peliosis hepatis, hepatic vein thrombosis, veno-occlusive disease, preeclampsia, eclampsia, acute fatty liver of pregnancy, intrahepatic cholestasis of pregnancy, and hepatic tumors including nodular hyperplasias, adenomas, and carcinomas; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation; an autoimmune/inflammatory disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis,

atherosclerotic plaque rupture, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, degradation of articular cartilage, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, bone resorption, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, age-related macular degeneration, and sensorineural hearing loss; an epithelial disorder, such as dyshidrotic eczema, allergic contact dermatitis, keratosis pilaris, melasma, vitiligo, actinic keratosis, basal cell carcinoma, squamous cell carcinoma, seborrheic keratosis, folliculitis, herpes simplex, herpes zoster, varicella, candidiasis, dermatophytosis, scabies, insect bites, cherry angioma, keloid, dermatofibroma, acrochordons, urticaria, transient acantholytic dermatosis, xerosis, eczema, atopic dermatitis, contact dermatitis, hand eczema, nummular eczema, lichen simplex chronicus, asteatotic eczema, stasis dermatitis and stasis ulceration, seborrheic dermatitis, psoriasis, lichen planus, pityriasis rosea, impetigo, ecthyma, dermatophytosis, tinea versicolor, warts, acne vulgaris, acne rosacea, pemphigus vulgaris, pemphigus foliaceus, paraneoplastic pemphigus, bullous pemphigoid, herpes gestationis, dermatitis herpetiformis, linear IgA disease, epidermolysis bullosa acquisita, dermatomyositis, lupus

erythematosus, scleroderma and morphea, erythroderma, alopecia, figurate skin lesions, telangiectasias, hypopigmentation, hyperpigmentation, vesicles/bullae, exanthems, cutaneous drug reactions, papulonodular skin lesions, chronic non-healing wounds, photosensitivity diseases, epidermolysis bullosa simplex, epidermolytic hyperkeratosis, epidermolytic and nonepidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, ichthyosis exfoliativa, keratosis palmaris et plantaris, keratosis palmoplantaris, palmoplantar keratoderma, keratosis punctata, Meesmann's corneal dystrophy, pachyonychia congenita, white sponge nevus, steatocystoma multiplex, epidermal nevi/epidermolytic hyperkeratosis type, monilethrix, trichothiodystrophy, chronic hepatitis/cryptogenic cirrhosis, and colorectal hyperplasia; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; and a reproductive disorder, such as infertility, including tubal disease, ovulatory defects, and endometriosis, a disorder of prolactin production, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, an endometrial or ovarian tumor, a uterine fibroid, autoimmune disorders, an ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea: a disruption of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia. The polynucleotide sequences encoding PRTS may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies;

in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PRTS expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PRTS may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PRTS may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding PRTS in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PRTS, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PRTS, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PRTS may involve the use of PCR. These oligomers may be chemically synthesized, generated

enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PRTS, or a fragment of a polynucleotide complementary to the polynucleotide encoding PRTS, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PRTS are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PRTS include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene

function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PRTS, fragments of PRTS, or antibodies specific for PRTS may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is

not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the

polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PRTS to quantify the levels of PRTS expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci.

USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PRTS may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PRTS on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of

the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, PRTS, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PRTS and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with PRTS, or fragments thereof, and washed. Bound PRTS is then detected by methods well known in the art. Purified PRTS can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding PRTS specifically compete with a test compound for binding PRTS. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PRTS.

In additional embodiments, the nucleotide sequences which encode PRTS may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/190,708, U.S. Ser. No.60/193,182, U.S. Ser. No.60/197,086, U.S. Ser. No.60/199,022, and U.S. Ser. No.60/200,227, are hereby expressly incorporated by reference herein.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized

and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA

assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:12-22. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative proteases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode proteases, the encoded polypeptides were analyzed by querying against PFAM

models for proteases. Potential proteases were also identified by homology to Incyte cDNA sequences that had been annotated as proteases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of PRTS Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:12-22 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:12-22 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs

from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PRTS are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma,

cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding PRTS. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of PRTS Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μl to 10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and

sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:12-22 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature

under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/µl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/µl RNase inhibitor, 500 µM dATP, 500 µM dGTP, 500 µM dTTP, 40 µM dCTP, 40 µM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)$^+$ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered

with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and

measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the PRTS-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PRTS. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PRTS. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PRTS-encoding transcript.

XII. Expression of PRTS

Expression and purification of PRTS is achieved using bacterial or virus-based expression systems. For expression of PRTS in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express PRTS upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of PRTS in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PRTS by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, PRTS is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step,

affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PRTS at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified PRTS obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, XVIII and XIX where applicable.

XIII. Functional Assays

PRTS function is assessed by expressing the sequences encoding PRTS at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PRTS on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding PRTS and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human

immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding PRTS and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of PRTS Specific Antibodies

PRTS substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PRTS amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-PRTS activity by, for example, binding the peptide or PRTS to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring PRTS Using Specific Antibodies

Naturally occurring or recombinant PRTS is substantially purified by immunoaffinity chromatography using antibodies specific for PRTS. An immunoaffinity column is constructed by covalently coupling anti-PRTS antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing PRTS are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PRTS (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PRTS binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PRTS is collected.

XVI. Identification of Molecules Which Interact with PRTS

PRTS, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PRTS, washed, and any wells with labeled PRTS complex are assayed. Data obtained using different concentrations of PRTS are used to calculate values for the number, affinity, and association of PRTS with the candidate molecules.

Alternatively, molecules interacting with PRTS are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

PRTS may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVII. Demonstration of PRTS Activity

Protease activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York NY, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases, or metalloproteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (carboxypeptidases A and B, procollagen C-proteinase). Commonly used chromogens are 2-naphthylamine, 4-nitroaniline, and furylacrylic acid. Assays are performed at ambient temperature and contain an aliquot of the enzyme and the appropriate substrate in a suitable buffer. Reactions are carried out in an optical cuvette, and the increase/decrease in absorbance of the chromogen released during hydrolysis of the peptide substrate is measured. The change in absorbance is proportional to the enzyme activity in the assay.

An alternate assay for ubiquitin hydrolase activity measures the hydrolysis of a ubiquitin precursor. The assay is performed at ambient temperature and contains an aliquot of PRTS and the appropriate substrate in a suitable buffer. Chemically synthesized human ubiquitin-valine may be used as substrate. Cleavage of the C-terminal valine residue from the substrate is monitored by capillary electrophoresis (Franklin, K. et al. (1997) Anal. Biochem. 247:305-309).

In the alternative, an assay for protease activity takes advantage of fluorescence resonance energy transfer (FRET) that occurs when one donor and one acceptor fluorophore with an appropriate spectral overlap are in close proximity. A flexible peptide linker containing a cleavage site specific for PRTS is fused between a red-shifted variant (RSGFP4) and a blue variant (BFP5) of Green

Fluorescent Protein. This fusion protein has spectral properties that suggest energy transfer is occurring from BFP5 to RSGFP4. When the fusion protein is incubated with PRTS, the substrate is cleaved, and the two fluorescent proteins dissociate. This is accompanied by a marked decrease in energy transfer which is quantified by comparing the emission spectra before and after the addition of PRTS (Mitra, R.D. et al. (1996) Gene 173:13-17). This assay can also be performed in living cells. In this case the fluorescent substrate protein is expressed constitutively in cells and PRTS is introduced on an inducible vector so that FRET can be monitored in the presence and absence of PRTS (Sagot, I. et al. (1999) FEBS Lett. 447:53-57).

XVIII. Identification of PRTS Substrates

Phage display libraries can be used to identify optimal substrate sequences for PRTS. A random hexamer followed by a linker and a known antibody epitope is cloned as an N-terminal extension of gene III in a filamentous phage library. Gene III codes for a coat protein, and the epitope will be displayed on the surface of each phage particle. The library is incubated with PRTS under proteolytic conditions so that the epitope will be removed if the hexamer codes for a PRTS cleavage site. An antibody that recognizes the epitope is added along with immobilized protein A. Uncleaved phage, which still bear the epitope, are removed by centrifugation. Phage in the supernatant are then amplified and undergo several more rounds of screening. Individual phage clones are then isolated and sequenced. Reaction kinetics for these peptide substrates can be studied using an assay in Example XVII, and an optimal cleavage sequence can be derived (Ke, S.H. et al. (1997) J. Biol. Chem. 272:16603-16609).

To screen for <u>in vivo</u> PRTS substrates, this method can be expanded to screen a cDNA expression library displayed on the surface of phage particles (T7SELECT 10-3 Phage display vector, Novagen, Madison WI) or yeast cells (pYD1 yeast display vector kit, Invitrogen, Carlsbad CA). In this case, entire cDNAs are fused between Gene III and the appropriate epitope.

XIX. Identification of PRTS Inhibitors

Compounds to be tested are arrayed in the wells of a multi-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. PRTS activity is measured for each well and the ability of each compound to inhibit PRTS activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance PRTS activity.

In the alternative, phage display libraries can be used to screen for peptide PRTS inhibitors. Candidates are found among peptides which bind tightly to a protease. In this case, multi-well plate wells are coated with PRTS and incubated with a random peptide phage display library or a cyclic peptide library (Koivunen, E. et al. (1999) Nat. Biotechnol. 17:768-774). Unbound phage are washed

away and selected phage amplified and rescreened for several more rounds. Candidates are tested for PRTS inhibitory activity using an assay described in Example XVII.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
3741323	1	3741323CD1	12	3741323CB1
5291211	2	5291211CD1	13	5291211CB1
7472661	3	7472661CD1	14	7472661CB1
1795307	4	1795307CD1	15	1795307CB1
7472685	5	7472685CD1	16	7472685CB1
2777676	6	2777676CD1	17	2777676CB1
7473620	7	7473620CD1	18	7473620CB1
3405540	8	3405540CD1	19	3405540CB1
4069067	9	4069067CD1	20	4069067CB1
4317947	10	4317947CD1	21	4317947CB1
7473303	11	7473303CD1	22	7473303CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	3741323	g439713	1.50E-271	precursor of P100 serine protease of Ra-reactive factor [Homo sapiens] Takada,F. et al. (1993) Biochem. Biophys. Res. Commun. 196:1003-1009
2	5291211	g6572225	7.40E-70	dJ408N23.5 (novel protein simialr to aminopeptidase P [homo sapiens]
3	7472661	g1061159	9.40E-270	testicular metalloprotease-like, Disintegrin-like cysteine rich protein Iva [Macaca fascicularis]
4	1795307	g1816586	1.50E-192	LON1 protease [Zea mays] Fu,G.K. and D.M. Markovitz (1998) Biochemistry 17:1905-1909; Meinhardt, A. et al. (1999) Hum. Reprod. Update 5:108-119
5	7472685	g3687497	1.90E-25	putative mitochondrial membrane protease subunit 2 [S. pombe]
6	2777676	g2970661	5.9e-99	Calpain isoform Lp85 [Rattus norvegicus]
7	7473620	g4336196	1.00E-154	carboxypeptidase A1 precursor [Sus scrofa]
8	3405540	g6137097	3.80E-82	serine protease DESC1[Homo sapiens]
9	4069067	g2465568	0	ADAM7 [Mus musculus] Cornwall,G.A. and Hsia,N. (1997) ADAM7, a member of the ADAM (a disintegrin and metalloprotease) gene family is specifically expressed in the mouse anterior pituitary and epididymis Endocrinology 138:4262-4272
10	4317947	g1573829	8.60E-44	aminopeptidase P (pepP) [Haemophilus influenzae Rd]
11	7473303	g2358070	2.40E-87	trypsinogen 1 [Mus musculus]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	3741323CD1	728	T506 S42 T51 T96 S123 T129 T198 S396 T601 T612 T632 S642 S720 T117 S167 T180 T191 S293 S356 T361 S381 T388 T472 S501 S675 Y214 Y619	N49 N178 N385 N407 N533 N599 N640	Signal_cleavage: M1-A17	SPSCAN
					Signal peptide: M1-A19	HMMER
					Trypsin: I450-V711	HMMER-PFAM
					Serine proteases, trypsin family, active sites (trypsin_his.prf): P476-A531	PROFILESCAN
					Serine proteases, trypsin family, active sites (trypsin_ser.prf): V643-C692	PROFILESCAN
					Serine proteases, trypsin family, histidine proteins: BL00134A:G482-V498 BL00134B:D658-Q681 BL00134C:Y698-V711	BLIMPS_BLOCKS
					Chymotrypsin serine protease: PR00722A:S483-V498 PR00722B:N549-V563 PR00722C:K657-V669	BLIMPS-PRINTS
					Serine protease precursor: PD000046: I450-V711	BLAST-PRODOM
					C1R/C1S Repeat: DM00162|P48740|13-138: S13-D139	BLAST-DOMO
					Trypsin_Ser: D658-F668	MOTIFS
					CUB domain: H20-Y135, C185-Y294	HMMER-PFAM
2	5291211CD1	211	S27 S146 S32 S68 S91 S168 T177 T184		signal_cleavage: M1-V17	SPSCAN
					AMINOPEPTIDASE HYDROLASE P AMINOACYLPROLINE MANGANESE ZINC PD011173: L67-P208	BLAST-PRODOM
					AMINOPEPTIDASE P AND PROLINE DIPEPTIDASE DM03215|P44881|7-161: S108-N198	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7472661CD1	559	Y143 T49 T389 T9 S39 S151 T170 S201 T284 T350 S378 T385 S484 S526 T90 T91 S113 T170 S235 T280 S363	N165 N187 N222 N348 N463 N468	PROTEIN TRANSMEMBRANE FERTILIN ALPHA CELLULAR DISINTEGRIN INTEGRIN METALLOPROTEASE SPERM CYSTEINE RICH PD004807: C403-C467	BLAST-PRODOM
					do ZINC; NEUTRAL; METALLOPEPTIDASE; HEMORRHAGIC; DM00533|S59854|14-197: L14-S195	BLAST-DOMO
					signal_peptide: M1-C28	HMMER
					signal_cleavage: M1-G31	SPSCAN
					Disintegrin signature PR00289: E317-N329	BLIMPS-PRINTS
					transmembrane domain: M535-L552	HMMER
					Reprolysin family propeptide Pep_M12B_propep: H75-M190	HMMER-PFAM
					INTEGRIN CELLULAR DISINTEGRIN TESTICULAR METALLOPROTEASELIKE DISINTEGRINLIKE CYSTEINERICH PROTEIN RMDC4A RMDC4B PD013512: S158-R307	BLAST-PRODOM
4	1795307CD1	852	Y344 T252 T31 S194 S206 T257 S281 S303 S325 T514 T547 T595 S612 S675 S825 T28 S35 T136 S194 T231 T327 T595 S648 S757 T772 T840	N323 N470 N492	ENDOPEPTIDASE LA (LON) SERINE PROTEASE (S16) SIGNATURE PR00830A: G375-R394, P654-M670, V737-F756, M767-K786, A790-L808	BLIMPS-PRINTS
					PROTEASE ATPBINDING HYDROLASE SERINE ATPDEPENDENT LA PROTEIN HOMOLOG MITOCHONDRIAL LON PD001169: A697-L831	BLAST-PRODOM
					AAA-PROTEIN FAMILY DM00024|P36774|338-504: A335-L501	BLAST-DOMO
					ATP/GTP binding site (P-loop):375-T382	MOTIFS
					LON Serine protease: D740-I748	MOTIFS
					signal_peptide: M1-R30	HMMER
					AAA domain: I370-D565	HMMER-PFAM
					ATP-dependent serine protease signature BL01046B: D331-P376, P377-S420, Q452-N492	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	7472685CD1	166	S73 S97 T9	N55 N122	SIGNAL PEPTIDASES I LYSINE DM01239|P28627|16-168: C21-I141	BLAST-DOMO
					Signal peptidase I signature Spase_I_3: V113-S126	MOTIFS
					Signal peptidase I signature Peptidase_S26: G38-P98	HMMER-PFAM
					Signal peptidases I serine protease signature BL00501B: Y61-D76, I81-K92, V108-R127	BLIMPS-BLOCKS
					BACTERIAL LEADER PEPTIDASE SIGNATURE PR00727A: Y29-I45, I81-I93, V108-R127	BLIMPS-PRINTS
6	2777676CD1	669	S58 S466 S81 S179 S208 T254 S316 T342 S374 S495 S539 T658 T23 T34 S415 S429 T462 S480 T616 T644	N372 N444 N647	CALPAIN CYSTEINE PROTEASE PR00704: Q19-A42, K63-I85, Q87-T103, Y122-V147, L153-L176, G178-L205, E309-C330	BLIMPS-PRINTS
					PROTEASE CALPAIN HYDROLASE SUBUNIT NEUTRAL THIOL LARGE CALCIUMACTIVATED PROTEINASE CANP PD001545: F35-S331	BLAST-PRODOM
					CALPAIN CATALYTIC DOMAIN DM01305|A48764|1-507: I14-N367	BLAST-DOMO
					Calpain family cysteine protease Peptidase_C2: F35-I333	HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7473620CD1	436	S124 S258 S14 T141 S147 T238 S299 T416 T426 Y358	N36 N171 N256 N281	Zinc carboxypeptidases, zinc-binding regions signatures: E302-L357	PROFILESCAN
					CARBOXYPEPTIDASE A METALLOPROTEASE PR00765: I165-L177, P187-I201, G267-K275, I320-Y333	BLIMPS-PRINTS
					CARBOXYPEPTIDASE PRECURSOR HYDROLASE ZINC ZYMOGEN SIGNAL B A 3DSTRUCTURE A1 PD005637: G33-E117	BLAST-PRODOM
					ZINC CARBOXYPEPTIDASES, ZINC-BINDING REGION 1 DM00683|P15085|112-418: R129-P435	BLAST-DOMO
					Carboxypept_Zn_1 P187-T209	MOTIFS
					Carboxypept_Zn_2 H323-Y333	MOTIFS
					signal_cleavage: M1-G33	SPSCAN
					Zinc carboxypeptidase Zn_carbOpept: Y139-E419	HMMER-PFAM
					Zinc carboxypeptidases BL00132:Y139-F179, P187-W200, Y217-K257, P261-K275, P287-H313, N315-L336, T372-G389	BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	3405540CD1	416	T57 S190 S281 S299 S386 S8 S126 S152 S216 T250 S405 T411	N72 N87 N107 N231 N261 N315 N336 N364	Signal peptide: M1-G29	SPSCAN
					Transmembrane domain: W12-L39	HMMER
					Trypsin domain: I185-I410	HMMER-PFAM
					Serine proteases, trypsin family signature BL00134: C210-C226; D360-G383; P397-I410	BLIMPS-BLOCKS
					Type I fibronectin domain BL01253: C210-A223; A359-Y372; W379-K413	BLIMPS-BLOCKS
					Kringle domains BL00021: C210-F227; I289-G310; P369-I410	BLIMPS-BLOCKS
					Serine proteases, trypsin family, active sites: M202-N255; V345-K393	PROFILESCAN
					Chymotrypsin serine protease signature PR00722: G211-C226; G266-V280; A359-A371	BLIMPS-PRINTS
					AIRWAY TRYPSINLIKE PROTEASE PD103718: M1-C170	BLAST-PRODOM
					TRYPSIN DM00018|A57014|45-284:I185-K413	BLAST-DOMO
					Trypsin active site motif: L221-C226	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	4069067CD1	776	T43 T46 T54 S124 S169 S186 T253 S262 T281 T441 T506 T562 S91 T176 S292 S354 T586 T757 S774 Y61 Y140 Y147 Y200	N84 N167 N174 N184 N584 N668 N772	Transmembrane domain: I669-R691 Reprolysin family propeptide: H73-E188 Reprolysin (M12B) family zinc metalloprotease domain: K199-P394 Disintegrin domain: D411-R486 Disintegrins proteins signature BL00427: C428-K482 Disintegrins signature: G419-D483 Disintegrin signature PR00289: C442-R461; E471-D483 EPIDIDYMAL APICAL PROTEIN IPRECURSOR SIGNAL A DISINTEGRIN METALLOPROTEASE DOMAIN ADAM PD019895: R691-S774 do ZINC; METALLOPEPTIDASE; NEUTRAL; ATROLYSIN; DM00368|S28258|192-394: I192-F395	HMMER HMMER-PFAM HMMER-PFAM HMMER-PFAM BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
10	4317947CD1	238	T110 T3 Y31	N39	Signal peptide: M1-G37 Metallopeptidase family M24 domain: M1-K225 Aminopeptidase P and proline dipeptidase signature BL00491: V34-H45, H151-D163, L173-I187, G202-S215 Methionine aminopeptidase signature PR00599: F142-G154; L171-P183 AMINOPEPTIDASE HYDROLASE METHIONINE PEPTIDASE PROTEIN COBALT M DIPEPTIDASE XPRO MAP PD000555: A30-L217 AMINOPEPTIDASE P AND PROLINE DIPEPTIDASE DM00816|P40051|237-484: E10-S215	SPSCAN HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	7473303CD1	250	T96 T109 T126 S219 Y116	N11 N165 N182	Signal peptide: M1-P20	HMMER, SPSCAN
					Trypsin domain: C41-I243	HMMER-PFAM
					Kringle domain BL00021: C41-N58; E202-I243	BLIMPS-BLOCKS
					Serine proteases, trypsin family signature BL00134:C41-C57, Q198-A221, V230-I243	BLIMPS-BLOCKS
					Type I fibronectin domain BL01253: C41-A54	BLIMPS-BLOCKS
					Serine proteases, trypsin family, active sites: K35-K77	PROFILESCAN
					Chymotrypsin serine protease signature PR00722: A42-C57; S97-A111	BLIMPS-PRINTS
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: C41-I243	BLAST-PRODOM
					TRYPSIN DM00018|S55065|26-244: S26-I247	BLAST-DOMO
					Trypsin active site motif: I52-C57	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
12	3741323CB1	3863	887-3422	3741323H1 (MENTNOT01)	1400	1684
				70885185V1	2434	2931
				542338T6 (OVARNOT02)	3227	3863
				6464949H1 (PLACFEB01)	2153	2861
				70886870V1	501	1232
				70888356V1	1	589
				779780R1 (MYOMNOT01)	2916	3485
				70887223V1	1634	2251
				70887431V1	667	1297
				70886621V1	1789	2426
				71134611V1	2811	3400
				71128957V1	1102	1675
13	5291211CB1	1129	1-640	70998523V1	542	1129
				1854879F6 (HNT3AZT01)	1	626
14	7472661CB1	2196	1-1933	GNN.g6693583_012	1	2196
15	1795307CB1	2837	1-2231	3369279H1 (CONNTUT04)	1097	1360
				1661442H1 (BRSTNOT09)	1524	1758
				4421842H1 (BRAPDIT01)	1776	2015
				1208682R6 (BRSTNOT02)	2435	2837
				1353572T6 (LATRTUT02)	551	1100
				6802184H1 (COLENOR03)	1	604
				3725924H1 (BRSTNOT23)	1841	2131
				4756064H1 (BRAHNOT01)	1588	1849
				3485842H1 (KIDNNOT31)	1310	1591
				2563947R6 (ADRETUT01)	2024	2553
				60105605D1	852	1303
				2507041T6 (CONUTUT02)	2170	2750
16	7472685CB1	716	1-104	6843137H1 (BRSTNON02)	1	639
				g2821051	368	716
				70197538V1	234	681

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
17	2777676CB1	2894	1-2302, 2865-2894	1283503T6 (COLNNOT16)	2210	2864
				7073586BV1	2369	2890
				7058868H1 (BRALNON02)	1	542
				7052256BV1	758	1279
				2727313F6 (OVARTUT05)	521	1060
				6828422H1 (SINTNOR01)	1071	1577
				7073452BV1	1683	2262
				3833188H1 (PANCNOT17)	1293	1586
				7073627BV1	1455	2064
				g3739336	1074	1444
18	7473620CB1	1444	1-769, 839-905	2554476T6 (THYMNOT03)	833	1049
				2725281T6 (OVARTUT05)	930	1437
				GNN.g5306288_002.edit	1	1311
19	3405540CB1	1872	129-1523	7084533BV1	1277	1872
				7084427BV1	237	897
				7601724H1 (ESOGTME01)	5	629
				3403665H1 (ESOGNOT03)	1	242
				7122443BV1	935	1426
				7122416BV1	774	1363
				GNN.g7321570.edit	1	2605
20	4069067CB1	2605	1-1406, 1594-2370	4069067H1 (SEMVNOT05)	1657	1925
				FL333966_00001	99	1591
21	4317947CB1	1896	1-569	2017544T6 (THP1NOT01)	1284	1896
				2017544F6 (THP1NOT01)	972	1629
				4317947H1 (BRADDIT02)	1	264
				g2112805	622	997
22	7473303CB1	997	1-57, 594-621	GBI.g1552494.raw.comp	323	908
				g2114700	1	572

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
12	3741323CB1	PROSNOT14
13	5291211CB1	HNT3AZT01
15	1795307CB1	PROSTUT05
16	7472685CB1	PENCNOT09
17	2777676CB1	LUNGDIS03
18	7473620CB1	OVARTUT05
19	3405540CB1	ESOGNOT03
20	4069067CB1	SEMVNOT05
21	4317947CB1	THP1NOT01

Table 6

Library	Vector	Library Description
ESOGNOT03	pINCY	Library was constructed using RNA isolated from esophageal tissue obtained from a 53-year-old Caucasian male during a partial esophagectomy, proximal gastrectomy, and regional lymph node biopsy. Patient history included membranous nephritis, hyperlipidemia, benign hypertension, and anxiety state. Previous surgeries included an adenotonsillectomy. Family history included atherosclerotic coronary artery disease, cirrhosis, an abdominal aortic aneurysm rupture, breast cancer , a myocardial infarction, atherosclerotic coronary artery disease, myocardial infarction, and atherosclerotic coronary artery disease.
HNT3AZT01	pINCY	Library was constructed using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated for three days with 0.35 micromolar 5-aza-2'-deoxycytidine (AZ).
LUNGDIS03	pINCY	Library was constructed using diseased lung tissue. 7.6 x 10e5 clones from a diseased lung tissue library were subjected to two rounds of subtraction hybridization with 5.1 million clones from a normal lung tissue library. The starting library for subtraction was constructed using polyA RNA isolated from diseased lung tissue. Patient history included idiopathic pulmonary disease. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) Nucleic Acids Res. 19:1954; and Bonaldo et al. Genome Res. (1996) 6:791.
OVARTUT05	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 62-year-old Caucasian female during a total abdominal hysterectomy, removal of the fallopian tubes and ovaries, exploratory laparotomy, regional lymph node excision, and dilation and curettage. Pathology indicated a grade 4 endometrioid carcinoma with extensive squamous differentiation, forming a solid mass in the right ovary. The uterine endometrium was inactive, the cervix showed mild chronic cervicitis, and focal endometriosis was observed in the posterior uterine serosa. Curettings indicated weakly proliferative endometrium with excessive stromal breakdown in the uterus, and a prior cervical biopsy indicated mild chronic cervicitis with a prominent nabothian cyst in the cervix. Patient history included longitudinal deficiency of the radioulna, osteoarthritis, thrombophlebitis, and abnormal blood chemistries. Family history included atherosclerotic coronary artery disease, pulmonary embolism, and cerebrovascular disease.

Table 6 (cont.)

Library	Vector	Library Description
PENCNOT09	pINCY	Library was constructed using RNA isolated from penis right corpora cavernosa tissue removed from a male.
PROSNOT14	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 60-year-old Caucasian male during radical prostatectomy and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). Patient history included a kidney cyst and hematuria. Family history included benign hypertension, cerebrovascular disease, and arteriosclerotic coronary artery disease.
PROSTUT05	PSPORT1	Library was constructed using RNA isolated from prostate tumor tissue removed from a 69-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenocarcinoma (Gleason grade 3+4). Adenofibromatous hyperplasia was also present. Family history included congestive heart failure, multiple myeloma, hyperlipidemia, and rheumatoid arthritis.
SEMVNOT05	pINCY	Library was constructed using RNA isolated from seminal vesicle tissue removed from a 67-year-old Caucasian male during radical prostatectomy. Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason grade 3 + 3.
THP1NOT01	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 (ATCC TIB 202) is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (re: Int. J. cancer (1980) 26:171).

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85.2444-2448, Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	*PFAM hits:* Probability value= 1.0E-3 or less *Signal peptide hits:* Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,

b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,

c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and

d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

2. An isolated polypeptide of claim 1, having a sequence selected from the group consisting of SEQ ID NO:1-11.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4, having a sequence selected from the group consisting of SEQ ID NO:12-22.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has the sequence selected from the group consisting of SEQ ID NO:1-11.

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,

b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,

c) a polynucleotide complementary to the polynucleotide of a),

d) a polynucleotide complementary to the polynucleotide of b), and

e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.

19. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional PRTS, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and

b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,

b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and

c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence of claim 12, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,

b) detecting altered expression of the target polynucleotide, and

c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method for assessing toxicity of a test compound, said method comprising:

a) treating a biological sample containing nucleic acids with the test compound;

b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof;

c) quantifying the amount of hybridization complex; and

d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

31. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

32. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

33. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

34. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

35. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

36. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

37. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

38. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

39. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

40. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

41. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

42. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

43. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

44. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

45. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

46. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

47. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

48. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

49. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:20.

50. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

51. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

52. A diagnostic test for a condition or disease associated with the expression of PRST in a biological sample comprising the steps of:

a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody: polypeptide complex; and

b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

53. The antibody of claim 11, wherein the antibody is:

a) a chimeric antibody;

b) a single chain antibody;

c) a Fab fragment;

d) a $F(ab')_2$ fragment; or

e) a humanized antibody.

54. A composition comprising an antibody of claim 11 and an acceptable excipient.

55. A method of diagnosing a condition or disease associated with the expression of PRST in a subject, comprising administering to said subject an effective amount of the composition of claim 54.

56. A composition of claim 54, wherein the antibody is labeled.

57. A method of diagnosing a condition or disease associated with the expression of PRST in a subject, comprising administering to said subject an effective amount of the composition of claim 56.

58. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11 comprising:

a) immunizing an animal with a polypeptide of SEQ ID NO:1-11 or an immunogenic fragment thereof under conditions to elicit an antibody response;

b) isolating antibodies from said animal; and

c) screening the isolated antibodies with the polypeptide thereby identifying a polyclonal antibody which binds specifically to a polypeptide of SEQ ID NO:1-11.

59. An antibody produced by a method of claim 58.

60. A composition comprising the antibody of claim 59 and a suitable carrier.

61. A method of making a monoclonal antibody with the specificity of the antibody of claim 11 comprising:

a) immunizing an animal with a polypeptide of SEQ ID NO:1-11 or an immunogenic fragment thereof under conditions to elicit an antibody response;

b) isolating antibody producing cells from the animal;

c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;

d) culturing the hybridoma cells; and

e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide of SEQ ID NO:1-11.

62. A monoclonal antibody produced by a method of claim 61.

63. A composition comprising the antibody of claim 62 and a suitable carrier.

64. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

65. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

66. A method for detecting a polypeptide of SEQ ID NO:1-11 in a sample comprising the steps of:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide of SEQ ID NO:1-11 in the sample.

67. A method of purifying a polypeptide of SEQ ID NO:1-11 from a sample, the method comprising:

a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and

b) separating the antibody from the sample and obtaining purified polypeptide of SEQ ID NO:1-11.

<110> INCYTE GENOMICS, INC.
YUE, Henry
LU, Dyung Aina M. Lu
POLICKY, Jennifer L.
DELEGEANE, Angelo M.
TRIBOULEY, Catherine M.
KHAN, Farrah A.
AU-YOUNG, Janice
BANDMAN, Olga
LAL, Preeti
BOROWSKY, Mark L.
GANDHI, Ameena R.
HILLMAN, Jennifer L.
TANG, Y. Tom
BURFORD, Neil
BAUGHN, Mariah R.
NGUYEN, Danniel B.
YAO, Monique G.
WALIA, Narinder K.
HE, Ann
HAFALIA, April
LU, Yan
PATTERSON, Chandra

<120> PROTEASES

<130> PI-0062 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/190,708; 60/193,182; 60/197,086; 60/199,022; 60/200,227
<151> 2000-03-17; 2000-03-30; 2000-04-14; 2000-04-20; 2000-04-28

<160> 22
<170> PERL Program

<210> 1
<211> 728
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3741323CD1

<400> 1
Met Arg Trp Leu Leu Leu Tyr Tyr Ala Leu Cys Phe Ser Leu Ser
1 5 10 15
Lys Ala Ser Ala His Thr Val Glu Leu Asn Asn Met Phe Gly Gln
20 25 30
Ile Gln Ser Pro Gly Tyr Pro Asp Ser Tyr Pro Ser Asp Ser Glu
35 40 45
Val Thr Trp Asn Ile Thr Val Pro Asp Gly Phe Arg Ile Lys Leu
50 55 60
Tyr Phe Met His Phe Asn Leu Glu Ser Ser Tyr Leu Cys Glu Tyr
65 70 75
Asp Tyr Val Lys Val Glu Thr Glu Asp Gln Val Leu Ala Thr Phe
80 85 90
Cys Gly Arg Glu Thr Thr Asp Thr Glu Gln Thr Pro Gly Gln Glu
95 100 105
Val Val Leu Ser Pro Gly Ser Phe Met Ser Ile Thr Phe Arg Ser
110 115 120
Asp Phe Ser Asn Glu Glu Arg Phe Thr Gly Phe Asp Ala His Tyr
125 130 135

Met Ala Val Asp Val Asp Glu Cys Lys Glu Arg Glu Asp Glu Glu
140 145 150
Leu Ser Cys Asp His Tyr Cys His Asn Tyr Ile Gly Gly Tyr Tyr
155 160 165
Cys Ser Cys Arg Phe Gly Tyr Ile Leu His Thr Asp Asn Arg Thr
170 175 180
Cys Arg Val Glu Cys Ser Asp Asn Leu Phe Thr Gln Arg Thr Gly
185 190 195
Val Ile Thr Ser Pro Asp Phe Pro Asn Pro Tyr Pro Lys Ser Ser
200 205 210
Glu Cys Leu Tyr Thr Ile Glu Leu Glu Glu Gly Phe Met Val Asn
215 220 225
Leu Gln Phe Glu Asp Ile Phe Asp Ile Glu Asp His Pro Glu Val
230 235 240
Pro Cys Pro Tyr Asp Tyr Ile Lys Ile Lys Val Gly Pro Lys Val
245 250 255
Leu Gly Pro Phe Cys Gly Glu Lys Ala Pro Glu Pro Ile Ser Thr
260 265 270
Gln Ser His Ser Val Leu Ile Leu Phe His Ser Asp Asn Ser Gly
275 280 285
Glu Asn Arg Gly Trp Arg Leu Ser Tyr Arg Ala Ala Gly Asn Glu
290 295 300
Cys Pro Glu Leu Gln Pro Pro Val His Gly Lys Ile Glu Pro Ser
305 310 315
Gln Ala Lys Tyr Phe Phe Lys Asp Gln Val Leu Val Ser Cys Asp
320 325 330
Thr Gly Tyr Lys Val Leu Lys Asp Asn Val Glu Met Asp Thr Phe
335 340 345
Gln Ile Glu Cys Leu Lys Asp Gly Thr Trp Ser Asn Lys Ile Pro
350 355 360
Thr Cys Lys Ile Val Asp Cys Arg Ala Pro Gly Glu Leu Glu His
365 370 375
Gly Leu Ile Thr Phe Ser Thr Arg Asn Asn Leu Thr Thr Tyr Lys
380 385 390
Ser Glu Ile Lys Tyr Ser Cys Gln Glu Pro Tyr Tyr Lys Met Leu
395 400 405
Asn Asn Asn Thr Gly Ile Tyr Thr Cys Ser Ala Gln Gly Val Trp
410 415 420
Met Asn Lys Val Leu Gly Arg Ser Leu Pro Thr Cys Leu Pro Glu
425 430 435
Cys Gly Gln Pro Ser Arg Ser Leu Pro Ser Leu Val Lys Arg Ile
440 445 450
Ile Gly Gly Arg Asn Ala Glu Pro Gly Leu Phe Pro Trp Gln Ala
455 460 465
Leu Ile Val Val Glu Asp Thr Ser Arg Val Pro Asn Asp Lys Trp
470 475 480
Phe Gly Ser Gly Ala Leu Leu Ser Ala Ser Trp Ile Leu Thr Ala
485 490 495
Ala His Val Leu Arg Ser Gln Arg Arg Asp Thr Thr Val Ile Pro
500 505 510
Val Ser Lys Glu His Val Thr Val Tyr Leu Gly Leu His Asp Val
515 520 525
Arg Asp Lys Ser Gly Ala Val Asn Ser Ser Ala Ala Arg Val Val
530 535 540
Leu His Pro Asp Phe Asn Ile Gln Asn Tyr Asn His Asp Ile Ala
545 550 555
Leu Val Gln Leu Gln Glu Pro Val Pro Leu Gly Pro His Val Met
560 565 570
Pro Val Cys Leu Pro Arg Leu Glu Pro Glu Gly Pro Ala Pro His
575 580 585
Met Leu Gly Leu Val Ala Gly Trp Gly Ile Ser Asn Pro Asn Val
590 595 600
Thr Val Asp Glu Ile Ile Ser Ser Gly Thr Arg Thr Leu Ser Asp
605 610 615
Val Leu Gln Tyr Val Lys Leu Pro Val Val Pro His Ala Glu Cys
620 625 630

```
Lys Thr Ser Tyr Glu Ser Arg Ser Gly Asn Tyr Ser Val Thr Glu
                635                 640                 645
Asn Met Phe Cys Ala Gly Tyr Tyr Glu Gly Gly Lys Asp Thr Cys
            650                 655                 660
Leu Gly Asp Ser Gly Gly Ala Phe Val Ile Phe Asp Asp Leu Ser
            665                 670                 675
Gln Arg Trp Val Val Gln Gly Leu Val Ser Trp Gly Gly Pro Glu
            680                 685                 690
Glu Cys Gly Ser Lys Gln Val Tyr Gly Val Tyr Thr Lys Val Ser
            695                 700                 705
Asn Tyr Val Gly Trp Val Trp Glu Gln Met Gly Leu Pro Gln Ser
            710                 715                 720
Val Val Glu Pro Gln Val Glu Arg
            725
```

<210> 2
<211> 211
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5291211CD1

```
<400> 2
Met Pro Trp Leu Leu Ser Ala Pro Lys Leu Val Pro Ala Val Ala
  1               5                  10                  15
Asn Val Arg Gly Leu Ser Gly Cys Met Leu Cys Ser Gln Arg Arg
             20                  25                  30
Tyr Ser Leu Gln Pro Val Pro Glu Arg Arg Ile Pro Asn Arg Tyr
             35                  40                  45
Leu Gly Gln Pro Ser Pro Phe Thr His Pro His Leu Leu Arg Pro
             50                  55                  60
Gly Glu Val Thr Pro Gly Leu Ser Gln Val Glu Tyr Ala Leu Arg
             65                  70                  75
Arg His Lys Leu Met Ser Leu Ile Gln Lys Glu Ala Gln Gly Gln
             80                  85                  90
Ser Gly Thr Asp Gln Thr Val Val Val Leu Ser Asn Pro Thr Tyr
             95                 100                 105
Tyr Met Ser Asn Asp Ile Pro Tyr Thr Phe His Gln Asp Asn Asn
            110                 115                 120
Phe Leu Tyr Leu Cys Gly Phe Gln Glu Pro Asp Ser Ile Leu Val
            125                 130                 135
Leu Gln Ser Leu Pro Gly Lys Gln Leu Pro Ser His Lys Ala Ile
            140                 145                 150
Leu Phe Val Pro Arg Arg Asp Pro Ser Arg Glu Leu Trp Asp Gly
            155                 160                 165
Pro Arg Ser Gly Thr Asp Gly Ala Ile Ala Leu Thr Gly Val Asp
            170                 175                 180
Glu Ala Tyr Thr Leu Glu Glu Phe Gln His Leu Leu Pro Lys Met
            185                 190                 195
Lys Gly Asn Lys Trp Glu Gln Lys Ser His Tyr Lys Pro Asp Trp
            200                 205                 210
Asp
```

<210> 3
<211> 559
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472661CD1

<400> 3

```
Met Arg Gln Ala Glu Ala Arg Val Thr Leu Arg Ala Pro Leu Leu
 1               5                  10                  15
Leu Leu Gly Leu Trp Val Leu Leu Thr Pro Val Arg Cys Ser Gln
            20                  25                  30
Gly His Pro Ser Trp His Tyr Ala Ser Ser Lys Val Val Ile Pro
        35                  40                  45
Arg Lys Glu Thr His His Gly Lys Asp Leu Gln Phe Leu Gly Trp
    50                  55                  60
Leu Ser Tyr Ser Leu His Phe Gly Gly Gln Arg His Ile Ile His
65                  70                  75
Met Arg Arg Lys His Leu Leu Trp Pro Arg His Leu Leu Val Thr
                80                  85                  90
Thr Gln Asp Asp Gln Gly Ala Leu Gln Met Asp Asp Pro Tyr Ile
            95                  100                 105
Pro Pro Asp Cys Tyr Tyr Leu Ser Tyr Leu Glu Glu Val Pro Leu
        110                 115                 120
Ser Met Val Thr Val Asp Met Cys Cys Gly Gly Leu Arg Gly Ile
    125                 130                 135
Met Lys Leu Asp Asp Leu Ala Tyr Glu Ile Lys Pro Leu Gln Asp
140                 145                 150
Ser Arg Arg Leu Glu His Val Ser Gln Ile Val Ala Glu Pro Asn
                155                 160                 165
Ala Thr Gly Pro Thr Phe Arg Asp Gly Asp Asn Glu Glu Thr Asn
            170                 175                 180
Pro Leu Phe Ser Glu Ala Asn Asp Ser Met Asn Pro Arg Ile Ser
        185                 190                 195
Asn Trp Leu Tyr Ser Ser His Arg Gly Asn Ile Lys Gly Tyr Val
    200                 205                 210
Gln Cys Ser Asn Ser Tyr Cys Arg Val Asp Asp Asn Ile Thr Thr
215                 220                 225
Cys Ser Lys Glu Val Val Gln Met Phe Ser Leu Ser Asp Ser Ile
                230                 235                 240
Val Gln Asn Ile Asp Leu Arg Tyr Tyr Ile Tyr Leu Leu Thr Ile
            245                 250                 255
Tyr Asn Asn Cys Asp Pro Ala Pro Val Asn Asp Tyr Arg Val Gln
        260                 265                 270
Ser Ala Met Phe Thr Tyr Phe Arg Thr Thr Phe Phe Asp Thr Phe
    275                 280                 285
Arg Val His Ser Pro Thr Leu Leu Ile Lys Glu Ala Pro His Glu
290                 295                 300
Cys Asn Tyr Glu Pro Gln Arg Pro Ile Gln Asn Ile Cys Asp Leu
305                 310                 315
Pro Glu Tyr Cys His Gly Thr Thr Val Thr Cys Pro Ala Asn Phe
                320                 325                 330
Tyr Met Gln Asp Gly Thr Pro Cys Thr Glu Glu Gly Tyr Cys Tyr
            335                 340                 345
His Gly Asn Cys Thr Asp Arg Asn Val Leu Cys Lys Val Ile Phe
        350                 355                 360
Gly Val Ser Ala Glu Glu Ala Pro Glu Val Cys Tyr Asp Ile Asn
    365                 370                 375
Leu Glu Ser Tyr Arg Phe Gly His Cys Thr Arg Arg Gln Thr Ala
380                 385                 390
Leu Asn Asn Gln Ala Cys Ala Gly Ile Asp Lys Phe Cys Gly Arg
                395                 400                 405
Leu Gln Cys Thr Ser Val Thr His Leu Pro Arg Leu Gln Glu His
            410                 415                 420
Val Ser Phe His His Ser Val Thr Gly Gly Phe Gln Cys Phe Gly
        425                 430                 435
Leu Asp Asp His Arg Ala Thr Asp Thr Thr Asp Val Gly Cys Val
    440                 445                 450
Ile Asp Gly Thr Pro Cys Val His Gly Asn Phe Cys Asn Asn Thr
455                 460                 465
Arg Cys Asn Ala Thr Ile Thr Ser Leu Gly Tyr Asp Cys Arg Pro
                470                 475                 480
Glu Lys Cys Ser His Arg Gly Val Cys Asn Asn Arg Arg Asn Cys
            485                 490                 495
```

```
His Cys His Ile Gly Trp Asp Pro Pro Leu Cys Leu Arg Arg Gly
                500                 505                 510
Ala Gly Gly Ser Val Asp Ser Gly Pro Pro Pro Lys Ile Thr Arg
                515                 520                 525
Ser Val Lys Gln Ser Gln Gln Ser Val Met Tyr Leu Arg Val Val
                530                 535                 540
Phe Gly Arg Ile Tyr Thr Phe Ile Ile Ala Leu Leu Phe Gly Met
                545                 550                 555
Ala Thr Asn Val
```

<210> 4
<211> 852
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1795307CD1

<400> 4

```
Met Ser Ser Val Ser Pro Ile Gln Ile Pro Ser Arg Leu Pro Leu
  1               5                  10                  15
Leu Leu Thr His Glu Gly Val Leu Leu Pro Gly Ser Thr Met Arg
             20                  25                  30
Thr Ser Val Asp Ser Ala Arg Asn Leu Gln Leu Val Arg Ser Arg
             35                  40                  45
Leu Leu Lys Gly Thr Ser Leu Gln Ser Thr Ile Leu Gly Val Ile
             50                  55                  60
Pro Asn Thr Pro Asp Pro Ala Ser Asp Ala Gln Asp Leu Pro Pro
             65                  70                  75
Leu His Arg Ile Gly Thr Ala Ala Leu Ala Val Gln Val Val Gly
             80                  85                  90
Ser Asn Trp Pro Lys Pro His Tyr Thr Leu Leu Ile Thr Gly Leu
             95                 100                 105
Cys Arg Phe Gln Ile Val Gln Val Leu Lys Glu Lys Pro Tyr Pro
            110                 115                 120
Ile Ala Glu Val Glu Gln Leu Asp Arg Leu Glu Glu Phe Pro Asn
            125                 130                 135
Thr Cys Lys Met Arg Glu Glu Leu Gly Glu Leu Ser Glu Gln Phe
            140                 145                 150
Tyr Lys Tyr Ala Val Gln Leu Val Glu Met Leu Asp Met Ser Val
            155                 160                 165
Pro Ala Val Ala Lys Leu Arg Arg Leu Leu Asp Ser Leu Pro Arg
            170                 175                 180
Glu Ala Leu Pro Asp Ile Leu Thr Ser Ile Ile Arg Thr Ser Asn
            185                 190                 195
Lys Glu Lys Leu Gln Ile Leu Asp Ala Val Ser Leu Glu Glu Arg
            200                 205                 210
Phe Lys Met Thr Ile Pro Leu Leu Val Arg Gln Ile Glu Gly Leu
            215                 220                 225
Lys Leu Leu Gln Lys Thr Arg Lys Pro Lys Gln Asp Asp Asp Lys
            230                 235                 240
Arg Val Ile Ala Ile Arg Pro Ile Arg Arg Ile Thr His Ile Ser
            245                 250                 255
Gly Thr Leu Glu Asp Glu Asp Glu Asp Glu Asp Asn Asp Asp Ile
            260                 265                 270
Val Met Leu Glu Lys Lys Ile Arg Thr Ser Ser Met Pro Glu Gln
            275                 280                 285
Ala His Lys Val Cys Val Lys Glu Ile Lys Arg Leu Lys Lys Met
            290                 295                 300
Pro Gln Ser Met Pro Glu Tyr Ala Leu Thr Arg Asn Tyr Leu Glu
            305                 310                 315
Leu Met Val Glu Leu Pro Trp Asn Lys Ser Thr Thr Asp Arg Leu
            320                 325                 330
Asp Ile Arg Ala Ala Arg Ile Leu Leu Asp Asn Asp His Tyr Ala
```

335 340 345
Met Glu Lys Leu Lys Lys Arg Val Leu Glu Tyr Leu Ala Val Arg
350 355 360
Gln Leu Lys Asn Asn Leu Lys Gly Pro Ile Leu Cys Phe Val Gly
365 370 375
Pro Pro Gly Val Gly Lys Thr Ser Val Gly Arg Ser Val Ala Lys
380 385 390
Thr Leu Gly Arg Glu Phe His Arg Ile Ala Leu Gly Gly Val Cys
395 400 405
Asp Gln Ser Asp Ile Arg Gly His Arg Arg Thr Tyr Val Gly Ser
410 415 420
Met Pro Gly Arg Ile Ile Asn Gly Leu Lys Thr Val Gly Val Asn
425 430 435
Asn Pro Val Phe Leu Leu Asp Glu Val Asp Lys Leu Gly Lys Ser
440 445 450
Leu Gln Gly Asp Pro Ala Ala Ala Leu Leu Glu Val Leu Asp Pro
455 460 465
Glu Gln Asn His Asn Phe Thr Asp His Tyr Leu Asn Val Ala Phe
470 475 480
Asp Leu Ser Gln Val Leu Phe Ile Ala Thr Ala Asn Thr Thr Ala
485 490 495
Thr Ile Pro Ala Ala Leu Leu Asp Arg Met Glu Ile Ile Gln Val
500 505 510
Pro Gly Tyr Thr Gln Glu Glu Lys Ile Glu Ile Ala His Arg His
515 520 525
Leu Ile Pro Lys Gln Leu Glu Gln His Gly Leu Thr Pro Gln Gln
530 535 540
Ile Gln Ile Pro Gln Val Thr Thr Leu Asp Ile Ile Thr Arg Tyr
545 550 555
Thr Arg Glu Ala Gly Val Arg Ser Leu Asp Arg Lys Leu Gly Ala
560 565 570
Ile Cys Arg Ala Val Ala Val Lys Val Ala Glu Gly Gln His Lys
575 580 585
Glu Ala Lys Leu Asp Arg Ser Asp Val Thr Glu Arg Glu Gly Cys
590 595 600
Arg Glu His Ile Leu Glu Asp Glu Lys Pro Glu Ser Ile Ser Asp
605 610 615
Thr Thr Asp Leu Ala Leu Pro Pro Glu Met Pro Ile Leu Ile Asp
620 625 630
Phe His Ala Leu Lys Asp Ile Leu Gly Pro Pro Met Tyr Glu Met
635 640 645
Glu Val Ser Gln Arg Leu Ser Gln Pro Gly Val Ala Ile Gly Leu
650 655 660
Ala Trp Thr Pro Leu Gly Gly Glu Ile Met Phe Val Glu Ala Ser
665 670 675
Arg Met Asp Gly Glu Gly Gln Leu Thr Leu Thr Gly Gln Leu Gly
680 685 690
Asp Val Met Lys Glu Ser Ala His Leu Ala Ile Ser Trp Leu Arg
695 700 705
Ser Asn Ala Lys Lys Tyr Gln Leu Thr Asn Ala Phe Gly Ser Phe
710 715 720
Asp Leu Leu Asp Asn Thr Asp Ile His Leu His Phe Pro Ala Gly
725 730 735
Ala Val Thr Lys Asp Gly Pro Ser Ala Gly Val Thr Ile Val Thr
740 745 750
Cys Leu Ala Ser Leu Phe Ser Gly Arg Leu Val Arg Ser Asp Val
755 760 765
Ala Met Thr Gly Glu Ile Thr Leu Arg Gly Leu Val Leu Pro Val
770 775 780
Gly Gly Ile Lys Asp Lys Val Leu Ala Ala His Arg Ala Gly Leu
785 790 795
Lys Gln Val Ile Ile Pro Arg Arg Asn Glu Lys Asp Leu Glu Gly
800 805 810
Ile Pro Gly Asn Val Arg Gln Asp Leu Ser Phe Val Thr Ala Ser
815 820 825
Cys Leu Asp Glu Val Leu Asn Ala Ala Phe Asp Gly Gly Phe Thr

```
                830                 835                 840
Val Lys Thr Arg Pro Gly Leu Leu Asn Ser Lys Leu
                845                 850
```

<210> 5
<211> 166
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472685CD1

<400> 5
```
Met Leu Arg Gly Val Leu Gly Lys Thr Phe Arg Leu Val Gly Tyr
 1               5                  10                  15
Thr Ile Gln Tyr Gly Cys Ile Ala His Cys Ala Phe Glu Tyr Val
            20                  25                  30
Gly Gly Val Val Met Cys Ser Gly Pro Ser Met Glu Pro Thr Ile
        35                  40                  45
Gln Asn Ser Asp Ile Val Phe Ala Glu Asn Leu Ser Arg His Phe
    50                  55                  60
Tyr Gly Ile Gln Arg Gly Asp Ile Val Ile Ala Lys Ser Pro Ser
65                  70                  75
Asp Pro Lys Ser Asn Ile Cys Lys Arg Val Ile Gly Leu Glu Gly
                80                  85                  90
Asp Lys Ile Leu Thr Thr Ser Pro Ser Asp Phe Phe Lys Ser His
            95                  100                 105
Ser Tyr Val Pro Met Gly His Val Trp Leu Glu Gly Asp Asn Leu
        110                 115                 120
Gln Asn Ser Thr Asp Ser Arg Cys Tyr Gly Pro Ile Pro Tyr Gly
    125                 130                 135
Leu Ile Arg Gly Arg Ile Phe Phe Lys Ile Trp Pro Leu Ser Asp
140                 145                 150
Phe Gly Phe Leu Arg Ala Ser Pro Asn Gly His Arg Phe Ser Asp
155                 160                 165

Asp
```

<210> 6
<211> 669
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2777676CD1

<400> 6
```
Met Ala Tyr Tyr Gln Glu Pro Ser Val Glu Thr Ser Ile Ile Lys
 1               5                  10                  15
Phe Lys Asp Gln Asp Phe Thr Thr Leu Arg Asp His Cys Leu Ser
            20                  25                  30
Met Gly Arg Thr Phe Lys Asp Glu Thr Phe Pro Ala Ala Asp Ser
        35                  40                  45
Ser Ile Gly Gln Lys Leu Leu Gln Glu Lys Arg Leu Ser Asn Val
    50                  55                  60
Ile Trp Lys Arg Pro Gln Asp Leu Pro Gly Gly Pro Pro His Phe
65                  70                  75
Ile Leu Asp Asp Ile Ser Arg Phe Asp Ile Gln Gln Gly Gly Ala
                80                  85                  90
Ala Asp Cys Trp Phe Leu Ala Ala Leu Gly Ser Leu Thr Gln Asn
            95                  100                 105
Pro Gln Tyr Arg Gln Lys Ile Leu Met Val Gln Ser Phe Ser His
        110                 115                 120
Gln Tyr Ala Gly Ile Phe Arg Phe Arg Phe Trp Gln Cys Gly Gln
```

```
                125                 130                 135
Trp Val Glu Val Val Ile Asp Asp Arg Leu Pro Val Gln Gly Asp
            140                 145                 150
Lys Cys Leu Phe Val Arg Pro Arg His Gln Asn Gln Glu Phe Trp
        155                 160                 165
Pro Cys Leu Leu Glu Lys Ala Tyr Ala Lys Leu Leu Gly Ser Tyr
    170                 175                 180
Ser Asp Leu His Tyr Gly Phe Leu Glu Asp Ala Leu Val Asp Leu
185                 190                 195
Thr Gly Gly Val Ile Thr Asn Ile His Leu His Ser Ser Pro Val
            200                 205                 210
Asp Leu Val Lys Ala Val Lys Thr Ala Thr Lys Ala Gly Ser Leu
        215                 220                 225
Ile Thr Cys Ala Thr Pro Ser Gly Pro Thr Asp Thr Ala Gln Ala
    230                 235                 240
Met Glu Asn Gly Leu Val Ser Leu His Ala Tyr Thr Val Thr Gly
245                 250                 255
Ala Glu Gln Ile Gln Tyr Arg Arg Gly Trp Glu Glu Ile Ile Ser
            260                 265                 270
Leu Trp Asn Pro Trp Gly Trp Gly Glu Thr Glu Trp Arg Gly Arg
        275                 280                 285
Trp Ser Asp Gly Ser Gln Glu Trp Glu Glu Thr Cys Asp Pro Arg
    290                 295                 300
Lys Ser Gln Leu His Lys Lys Arg Glu Asp Gly Glu Phe Trp Met
305                 310                 315
Ser Cys Gln Asp Phe Gln Gln Lys Phe Ile Ala Met Phe Ile Cys
            320                 325                 330
Ser Glu Ile Pro Ile Thr Leu Asp His Gly Asn Thr Leu His Glu
        335                 340                 345
Gly Trp Ser Gln Ile Met Phe Arg Lys Gln Val Ile Leu Gly Asn
    350                 355                 360
Thr Ala Gly Gly Pro Arg Asn Asp Ala Gln Phe Asn Phe Ser Val
365                 370                 375
Gln Glu Pro Met Glu Gly Thr Asn Val Val Val Cys Val Thr Val
            380                 385                 390
Ala Val Thr Pro Ser Asn Leu Lys Ala Glu Asp Ala Lys Phe Pro
        395                 400                 405
Leu Asp Phe Gln Val Ile Leu Ala Gly Ser Gln Arg Phe Arg Glu
    410                 415                 420
Lys Phe Pro Pro Val Phe Phe Ser Ser Phe Arg Asn Thr Val Gln
425                 430                 435
Ser Ser Asn Asn Lys Phe Arg Arg Asn Phe Thr Met Thr Tyr His
            440                 445                 450
Leu Ser Pro Gly Asn Tyr Val Val Val Ala Gln Thr Arg Arg Lys
        455                 460                 465
Ser Ala Glu Phe Leu Leu Arg Ile Phe Leu Lys Met Pro Asp Ser
    470                 475                 480
Asp Arg His Leu Ser Ser His Phe Asn Leu Arg Met Lys Gly Ser
485                 490                 495
Pro Ser Glu His Gly Ser Gln Gln Ser Ile Phe Asn Arg Tyr Ala
            500                 505                 510
Gln Gln Arg Leu Asp Ile Asp Ala Thr Gln Leu Gln Gly Leu Leu
        515                 520                 525
Asn Gln Glu Leu Leu Thr Gly Pro Pro Gly Asp Met Phe Ser Leu
    530                 535                 540
Asp Glu Cys Arg Ser Leu Val Ala Leu Met Glu Leu Lys Val Asn
545                 550                 555
Gly Arg Leu Asp Gln Glu Glu Phe Ala Arg Leu Trp Lys Arg Leu
            560                 565                 570
Val His Tyr Gln His Val Phe Gln Lys Val Gln Thr Ser Pro Gly
        575                 580                 585
Val Leu Leu Ser Ser Asp Leu Trp Lys Ala Ile Glu Asn Thr Asp
    590                 595                 600
Phe Leu Arg Gly Ile Phe Ile Ser Arg Glu Leu Leu His Leu Val
605                 610                 615
Thr Leu Arg Tyr Ser Asp Ser Val Gly Arg Val Ser Phe Pro Ser
```

620 625 630
Leu Val Cys Phe Leu Met Arg Leu Glu Ala Met Ala Lys Thr Phe
635 640 645
Arg Asn Leu Ser Lys Asp Gly Lys Gly Leu Tyr Leu Thr Glu Met
650 655 660
Glu Trp Met Ser Leu Val Met Tyr Asn
665

<210> 7
<211> 436
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473620CD1

<400> 7
Met Gln Gly Thr Pro Gly Gly Gly Thr Arg Pro Gly Pro Ser Pro
1 5 10 15
Val Asp Arg Arg Thr Leu Leu Val Phe Ser Phe Ile Leu Ala Ala
20 25 30
Ala Leu Gly Gln Met Asn Phe Thr Gly Asp Gln Val Leu Arg Val
35 40 45
Leu Ala Lys Asp Glu Lys Gln Leu Ser Leu Leu Gly Asp Leu Glu
50 55 60
Gly Leu Lys Pro Gln Lys Val Asp Phe Trp Arg Gly Pro Ala Arg
65 70 75
Pro Ser Leu Pro Val Asp Met Arg Val Pro Phe Ser Glu Leu Lys
80 85 90
Asp Ile Lys Ala Tyr Leu Glu Ser His Gly Leu Ala Tyr Ser Ile
95 100 105
Met Ile Lys Asp Ile Gln Val Leu Leu Asp Glu Glu Arg Gln Ala
110 115 120
Met Ala Lys Ser Arg Arg Leu Glu Arg Ser Thr Asn Ser Phe Ser
125 130 135
Tyr Ser Ser Tyr His Thr Leu Glu Glu Ile Tyr Ser Trp Ile Asp
140 145 150
Asn Phe Val Met Glu His Ser Asp Ile Val Ser Lys Ile Gln Ile
155 160 165
Gly Asn Ser Phe Glu Asn Gln Ser Ile Leu Val Leu Lys Phe Ser
170 175 180
Thr Gly Gly Ser Arg His Pro Ala Ile Trp Ile Asp Thr Gly Ile
185 190 195
His Ser Arg Glu Trp Ile Thr His Ala Thr Gly Ile Trp Thr Ala
200 205 210
Asn Lys Ile Val Ser Asp Tyr Gly Lys Asp Arg Val Leu Thr Asp
215 220 225
Ile Leu Asn Ala Met Asp Ile Phe Ile Glu Leu Val Thr Asn Pro
230 235 240
Asp Gly Phe Ala Phe Thr His Ser Met Asn Arg Leu Trp Arg Lys
245 250 255
Asn Lys Ser Ile Arg Pro Gly Ile Phe Cys Ile Gly Val Asp Leu
260 265 270
Asn Arg Asn Trp Lys Ser Gly Phe Gly Gly Asn Gly Ser Asn Ser
275 280 285
Asn Pro Cys Ser Glu Thr Tyr His Gly Pro Ser Pro Gln Ser Glu
290 295 300
Pro Glu Val Ala Ala Ile Val Asn Phe Ile Thr Ala His Gly Asn
305 310 315
Phe Lys Ala Leu Ile Ser Ile His Ser Tyr Ser Gln Met Leu Met
320 325 330
Tyr Pro Tyr Gly Arg Leu Leu Glu Pro Val Ser Asn Gln Arg Glu
335 340 345
Leu Tyr Asp Leu Ala Lys Asp Ala Val Glu Ala Leu Tyr Lys Val
350 355 360

His Gly Ile Glu Tyr Ile Phe Gly Ser Ile Ser Thr Thr Leu Tyr
365 370 375
Val Ala Ser Gly Ile Thr Val Asp Trp Ala Tyr Asp Ser Gly Ile
380 385 390
Lys Tyr Ala Phe Ser Phe Glu Leu Arg Asp Thr Gly Gln Tyr Gly
395 400 405
Phe Leu Leu Pro Ala Thr Gln Ile Ile Pro Thr Ala Gln Glu Thr
410 415 420
Trp Met Ala Leu Arg Thr Ile Met Glu His Thr Leu Asn His Pro
425 430 435
Tyr

<210> 8
<211> 416
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3405540CD1

<400> 8
Met Tyr Arg His Gly Ile Ser Ser Gln Arg Ser Trp Pro Leu Trp
1 5 10 15
Thr Thr Ile Phe Ile Phe Leu Gly Val Ala Ala Ile Leu Gly Val
20 25 30
Thr Ile Gly Leu Leu Val His Phe Leu Ala Val Glu Lys Thr Tyr
35 40 45
Tyr Tyr Gln Gly Asp Phe His Ile Ser Gly Val Thr Tyr Asn Asp
50 55 60
Asn Cys Glu Asn Ala Ala Ser Gln Ala Ser Thr Asn Leu Ser Lys
65 70 75
Asp Ile Glu Thr Lys Met Leu Asn Ala Phe Gln Asn Ser Ser Ile
80 85 90
Tyr Lys Glu Tyr Val Lys Ser Glu Val Ile Lys Leu Leu Pro Asn
95 100 105
Ala Asn Gly Ser Asn Val Gln Leu Gln Leu Lys Phe Lys Phe Pro
110 115 120
Pro Ala Glu Gly Val Ser Met Arg Thr Lys Ile Lys Ala Lys Leu
125 130 135
His Gln Met Leu Lys Asn Asn Met Ala Ser Trp Asn Ala Val Pro
140 145 150
Ala Ser Ile Lys Leu Met Glu Ile Ser Lys Ala Ala Ser Glu Met
155 160 165
Leu Thr Asn Asn Cys Cys Gly Arg Gln Val Ala Asn Ser Ile Ile
170 175 180
Thr Gly Asn Lys Ile Val Asn Gly Lys Ser Ser Leu Glu Gly Ala
185 190 195
Trp Pro Trp Gln Ala Ser Met Gln Trp Lys Gly Arg His Tyr Cys
200 205 210
Gly Ala Ser Leu Ile Ser Ser Arg Trp Leu Leu Ser Ala Ala His
215 220 225
Cys Phe Ala Lys Lys Asn Asn Ser Lys Asp Trp Thr Val Asn Phe
230 235 240
Gly Val Val Val Asn Lys Pro Tyr Met Thr Arg Lys Val Gln Asn
245 250 255
Ile Ile Phe His Glu Asn Tyr Ser Ser Pro Gly Leu His Asp Asp
260 265 270
Ile Ala Leu Val Gln Leu Ala Glu Glu Val Ser Phe Thr Glu Tyr
275 280 285
Ile Arg Lys Ile Cys Leu Pro Glu Ala Lys Met Lys Leu Ser Glu
290 295 300
Asn Asp Asn Val Val Val Thr Gly Trp Gly Thr Leu Tyr Met Asn
305 310 315
Gly Ser Phe Pro Val Ile Leu Gln Glu Ala Phe Leu Lys Ile Ile

320 325 330
Asp Asn Lys Ile Cys Asn Ala Ser Tyr Ala Tyr Ser Gly Phe Val
335 340 345
Thr Asp Ser Met Leu Cys Ala Gly Phe Met Ser Gly Glu Ala Asp
350 355 360
Ala Cys Gln Asn Asp Ser Gly Gly Pro Leu Ala Tyr Pro Asp Ser
365 370 375
Arg Asn Ile Trp His Leu Val Gly Ile Val Ser Trp Gly Asp Gly
380 385 390
Cys Gly Lys Lys Asn Lys Pro Gly Val Tyr Thr Arg Val Thr Ser
395 400 405
Tyr Arg Asn Trp Ile Thr Ser Lys Thr Gly Leu
410 415

<210> 9
<211> 776
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4069067CD1

<400> 9
Met Leu Pro Gly Cys Ile Phe Leu Met Ile Leu Leu Ile Pro Gln
1 5 10 15
Val Lys Glu Lys Phe Ile Leu Gly Val Glu Gly Gln Gln Leu Val
20 25 30
Arg Pro Lys Lys Leu Pro Leu Ile Gln Lys Arg Asp Thr Gly His
35 40 45
Thr His Asp Asp Asp Ile Leu Lys Thr Tyr Glu Glu Glu Leu Leu
50 55 60
Tyr Glu Ile Lys Leu Asn Arg Lys Thr Leu Val Leu His Leu Leu
65 70 75
Arg Ser Arg Glu Phe Leu Gly Ser Asn Tyr Ser Glu Thr Phe Tyr
80 85 90
Ser Met Lys Gly Glu Ala Phe Thr Arg His Pro Gln Ile Met Asp
95 100 105
His Cys Tyr Tyr Lys Gly Asn Ile Leu Asn Glu Lys Asn Ser Val
110 115 120
Ala Ser Ile Ser Thr Cys Asp Gly Leu Arg Gly Phe Phe Arg Ile
125 130 135
Asn Asp Gln Arg Tyr Leu Ile Glu Pro Val Lys Tyr Ser Asp Glu
140 145 150
Gly Glu His Leu Val Phe Lys Tyr Asn Leu Arg Val Pro Tyr Gly
155 160 165
Ala Asn Tyr Ser Cys Thr Glu Leu Asn Phe Thr Arg Lys Thr Val
170 175 180
Pro Gly Asp Asn Glu Ser Glu Glu Asp Ser Lys Ile Lys Gly Ile
185 190 195
His Asp Glu Lys Tyr Val Glu Leu Phe Ile Val Ala Asp Asp Thr
200 205 210
Val Tyr Arg Arg Asn Gly His Pro His Asn Lys Leu Arg Asn Arg
215 220 225
Ile Trp Gly Met Val Asn Phe Val Asn Met Ile Tyr Lys Thr Leu
230 235 240
Asn Ile His Val Thr Leu Val Gly Ile Glu Ile Trp Thr His Glu
245 250 255
Asp Lys Ile Glu Leu Tyr Ser Asn Ile Glu Thr Thr Leu Leu Arg
260 265 270
Phe Ser Phe Trp Gln Glu Lys Ile Leu Lys Thr Arg Lys Asp Phe
275 280 285
Asp His Val Val Leu Leu Ser Gly Lys Trp Leu Tyr Ser His Val
290 295 300
Gln Gly Ile Ser Tyr Pro Gly Gly Met Cys Leu Pro Tyr Tyr Ser
305 310 315

Thr Ser Ile Ile Lys Asp Leu Leu Pro Asp Thr Asn Ile Ile Ala
320 325 330
Asn Arg Met Ala His Gln Leu Gly His Asn Leu Gly Met Gln His
335 340 345
Asp Glu Phe Pro Cys Thr Cys Pro Ser Gly Lys Cys Val Met Asp
350 355 360
Ser Asp Gly Ser Ile Pro Ala Leu Asp Leu Ser Lys Cys Arg Gln
365 370 375
Asn Gln Tyr His Gln Tyr Leu Lys Asp Tyr Lys Pro Thr Cys Met
380 385 390
Leu Asn Ile Pro Phe Pro Tyr Asn Phe His Asp Phe Gln Phe Cys
395 400 405
Gly Asn Lys Lys Leu Asp Glu Gly Glu Glu Cys Asp Cys Gly Pro
410 415 420
Ala Gln Glu Cys Thr Asn Pro Cys Cys Asp Ala His Thr Cys Val
425 430 435
Leu Lys Pro Gly Phe Thr Cys Ala Glu Gly Glu Cys Cys Glu Ser
440 445 450
Cys Gln Ile Lys Lys Ala Gly Ser Ile Cys Arg Pro Ala Lys Asp
455 460 465
Glu Cys Asp Phe Pro Glu Met Cys Thr Gly His Ser Pro Ala Cys
470 475 480
Pro Lys Asp Gln Phe Arg Val Asn Gly Phe Pro Cys Lys Asn Ser
485 490 495
Glu Gly Tyr Cys Phe Met Gly Lys Cys Pro Thr Arg Glu Asp Gln
500 505 510
Cys Ser Glu Leu Phe Asp Asp Glu Ala Ile Glu Ser His Asp Ile
515 520 525
Cys Tyr Lys Met Asn Thr Lys Gly Asn Lys Phe Gly Tyr Cys Lys
530 535 540
Asn Lys Glu Asn Arg Phe Leu Pro Cys Glu Glu Lys Asp Val Arg
545 550 555
Cys Gly Lys Ile Tyr Cys Thr Gly Gly Glu Leu Ser Ser Leu Leu
560 565 570
Gly Glu Asp Lys Thr Tyr His Leu Lys Asp Pro Gln Lys Asn Ala
575 580 585
Thr Val Lys Cys Lys Thr Ile Phe Leu Tyr His Asp Ser Thr Asp
590 595 600
Ile Gly Leu Val Ala Ser Gly Thr Lys Cys Gly Glu Gly Met Val
605 610 615
Cys Asn Asn Gly Glu Cys Leu Asn Met Glu Lys Val Tyr Ile Ser
620 625 630
Thr Asn Cys Pro Ser Gln Cys Asn Glu Asn Pro Val Asp Gly His
635 640 645
Gly Leu Gln Cys His Cys Glu Glu Gly Gln Ala Pro Val Ala Cys
650 655 660
Glu Glu Thr Leu His Val Thr Asn Ile Thr Ile Leu Val Val Val
665 670 675
Leu Val Leu Val Ile Val Gly Ile Gly Val Leu Ile Leu Leu Val
680 685 690
Arg Tyr Arg Lys Cys Ile Lys Leu Lys Gln Val Gln Ser Pro Pro
695 700 705
Thr Glu Thr Leu Gly Val Glu Asn Lys Gly Tyr Phe Gly Asp Glu
710 715 720
Gln Gln Ile Arg Thr Glu Pro Ile Leu Pro Glu Ile His Phe Leu
725 730 735
Asn Gln Arg Thr Pro Glu Ser Leu Glu Ser Leu Pro Thr Ser Phe
740 745 750
Ser Ser Pro His Tyr Ile Thr Leu Lys Pro Ala Ser Lys Asp Ser
755 760 765
Arg Gly Ile Ala Asp Pro Asn Gln Ser Ala Lys
770 775

<210> 10
<211> 238
<212> PRT

<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4317947CD1

<400> 10
Met Phe Thr Ser Lys Ala Pro Val Glu Glu Ala Phe Leu Tyr Ala
1 5 10 15
Lys Phe Glu Phe Glu Cys Arg Ala Arg Gly Ala Asp Ile Leu Ala
20 25 30
Tyr Pro Pro Val Val Ala Gly Gly Asn Arg Ser Asn Thr Leu His
35 40 45
Tyr Val Lys Asn Asn Gln Leu Ile Lys Asp Gly Glu Met Val Leu
50 55 60
Leu Asp Gly Gly Cys Glu Ser Ser Cys Tyr Val Ser Asp Ile Thr
65 70 75
Arg Thr Trp Pro Val Asn Gly Arg Phe Thr Ala Pro Gln Ala Glu
80 85 90
Leu Tyr Glu Ala Val Leu Glu Ile Gln Arg Asp Cys Leu Ala Leu
95 100 105
Cys Phe Pro Gly Thr Ser Leu Glu Asn Ile Tyr Ser Met Met Leu
110 115 120
Thr Leu Ile Gly Gln Lys Leu Lys Asp Leu Gly Ile Met Lys Asn
125 130 135
Ile Lys Glu Asn Asn Ala Phe Lys Ala Ala Arg Lys Tyr Cys Pro
140 145 150
His His Val Gly His Tyr Leu Gly Met Asp Val His Asp Thr Pro
155 160 165
Asp Met Pro Arg Ser Leu Pro Leu Gln Pro Gly Met Val Ile Thr
170 175 180
Ile Glu Pro Gly Ile Tyr Ile Pro Glu Asp Asp Lys Asp Ala Pro
185 190 195
Glu Lys Phe Arg Gly Leu Gly Val Arg Ile Glu Asp Asp Val Val
200 205 210
Val Thr Gln Asp Ser Pro Leu Ile Leu Ser Ala Asp Cys Pro Lys
215 220 225
Glu Met Asn Asp Ile Glu Gln Ile Cys Ser Gln Ala Ser
230 235

<210> 11
<211> 250
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473303CD1

<400> 11
Met Lys Phe Ile Leu Leu Trp Ala Leu Leu Asn Leu Thr Val Ala
1 5 10 15
Leu Ala Phe Asn Pro Asp Tyr Thr Val Ser Ser Thr Pro Pro Tyr
20 25 30
Leu Val Tyr Leu Lys Ser Asp Tyr Leu Pro Cys Ala Gly Val Leu
35 40 45
Ile His Pro Leu Trp Val Ile Thr Ala Ala His Cys Asn Leu Pro
50 55 60
Lys Leu Arg Val Ile Leu Gly Val Thr Ile Pro Ala Asp Ser Asn
65 70 75
Glu Lys His Leu Gln Val Ile Gly Tyr Glu Lys Met Ile His His
80 85 90
Pro His Phe Ser Val Thr Ser Ile Asp His Asp Ile Met Leu Ile
95 100 105
Lys Leu Lys Thr Glu Ala Glu Leu Asn Asp Tyr Val Lys Leu Ala
110 115 120

```
Asn Leu Pro Tyr Gln Thr Ile Ser Glu Asn Thr Met Cys Ser Val
                125                 130                 135
Ser Thr Trp Ser Tyr Asn Val Cys Asp Ile Gly Ser Leu Thr Ser
            140                 145                 150
Ile Phe Ser Leu Asp Lys Glu Pro Asp Ser Leu Gln Thr Val Asn
        155                 160                 165
Ile Ser Val Ile Ser Lys Pro Gln Cys Arg Asp Ala Tyr Lys Thr
    170                 175                 180
Tyr Asn Ile Thr Glu Asn Met Leu Cys Val Gly Ile Val Pro Gly
185                 190                 195
Arg Arg Gln Pro Cys Lys Glu Val Ser Ala Ala Pro Ala Ile Cys
            200                 205                 210
Asn Gly Met Leu Gln Gly Ile Leu Ser Phe Ala Asp Gly Cys Val
        215                 220                 225
Leu Arg Ala Asp Val Gly Ile Tyr Ala Lys Ile Phe Tyr Tyr Ile
    230                 235                 240
Pro Trp Ile Glu Asn Val Ile Gln Asn Asn
245                 250
```

<210> 12
<211> 3863
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3741323CB1

<400> 12

```
cggacaccac accgggcacg agctcacagg caagtcaagc tgggaggacc aaggccgggc 60
agccgggagc acccaaggca ggaaaatgag gtggctgctt ctctattatg ctctgtgctt 120
ctccctgtca aaggcttcag cccacaccgt ggagctaaac aatatgtttg gccagatcca 180
gtcgcctggt tatccagact cctatcccag tgattcagag gtgacttgga atatcactgt 240
cccagatggg tttcggatca agctttactt catgcacttc aacttggaat cctcctacct 300
ttgtgaatat gactatgtga aggtagaaac tgaggaccag gtgctggcaa ccttctgtgg 360
cagggagacc acagacacag agcagactcc cggccaggag gtggtcctct cccctggctc 420
cttcatgtcc atcactttcc ggtcagattt ctccaatgag gagcgtttca caggctttga 480
tgcccactac atggctgtgg atgtggacga gtgcaaggag agggaggacg aggagctgtc 540
ctgtgaccac tactgccaca actacattgg cggctactac tgctcctgcc gcttcggcta 600
catcctccac acagacaaca ggacctgccg agtggagtgc agtgacaacc tcttcactca 660
aaggactggg gtgatcacca gccctgactt cccaaaccct taccccaaga gctctgaatg 720
cctgtatacc atcgagctgg aggagggttt catggtcaac ctgcagtttg aggacatatt 780
tgacattgag gaccatcctg aggtgccctg cccctatgac tacatcaaga tcaaagttgg 840
tccaaaagtt ttggggcctt tctgtggaga gaaagcccca gaacccatca gcacccagag 900
ccacagtgtc ctgatcctgt tccatagtga caactcggga gagaaccggg gctggaggct 960
ctcatacagg gctgcaggaa atgagtgccc agagctacag cctcctgtcc atgggaaaat 1020
cgagccctcc caagccaagt atttcttcaa agaccaagtg ctcgtcagct gtgacacagg 1080
ctacaaagtg ctgaaggata atgtggagat ggacacattc cagattgagt gtctgaagga 1140
tgggacgtgg agtaacaaga ttcccacctg taaaattgta gactgtagag ccccaggaga 1200
gctggaacac gggctgatca ccttctctac aaggaacaac ctcaccacat acaagtctga 1260
gatcaaatac tcctgtcagg agccctatta caagatgctc aacaataaca caggtatata 1320
tacctgttct gcccaaggag tctggatgaa taaagtattg gggagaagcc tacccacctg 1380
ccttccagag tgtggtcagc cctcccgctc cctgccaagt ctggtcaaga ggatcattgg 1440
gggccgaaat gctgagcccg gcctcttccc gtggcaggcc ctgatagtgg tggaggacac 1500
ttcgagagtg ccaaatgaca agtggtttgg gagtggggcc ctgctctctg cgtcctggat 1560
cctcacagca gctcatgtgc tgcgctccca gcgtagagac accacggtga taccagtctc 1620
caaggagcat gtcaccgtct acctgggctt gcatgatgtg cgagacaaat cgggggcagt 1680
caacagctca gctgcccgag tggtgctcca cccagacttc aacatccaaa actacaacca 1740
cgatatagct ctggtgcagc tgcaggagcc tgtgcccctg ggaccccacg ttatgcctgt 1800
ctgcctgcca aggcttgagc ctgaaggccc ggccccccac atgctgggcc tggtggccgg 1860
ctggggcatc tccaatccca atgtgacagt ggatgagatc atcagcagtg gcacacggac 1920
cttgtcagat gtcctgcagt atgtcaagtt acccgtggtg cctcacgctg agtgcaaaac 1980
tagctatgag tcccgctcgg gcaattacag cgtcacggag aacatgttct gtgctggcta 2040
ctacgagggc ggcaaagaca cgtgccttgg agatagcggt ggggcctttg tcatctttga 2100
tgacttgagc cagcgctggg tggtgcaagg cctggtgtcc tgggggggac ctgaagaatg 2160
cggcagcaag caggtctatg gagtctacac aaaggtctcc aattacgtgg gctgggtgtg 2220
```

ggagcagatg ggcttaccac aaagtgttgt ggagccccag gtggaacggt gagctgactt 2280
acttcctcgg ggcctgcctc ccctgagcga agctacaccg cacttccgac agcacactcc 2340
acattactta tcagaccata tggaatggaa cacactgacc tagcggtggc ttctcctacc 2400
gagacagccc ccaggaccct gagaggcaga gtgtggtata gggaaaaggc tccaggcagg 2460
agacctgtgt tcctgagctt gtccaagtct ctttccctgt ctgggcctca ctctaccgag 2520
taatacaatg caggagctca accaaggcct ctgtgccaat cccagcactc ctttccaggc 2580
catgcttctt accccagtgg cctttattca ctcctgacca cttatcaaac ccatcggtcc 2640
tactgttggt ataactgagc ttggacctga ctattagaaa atggtttcta acattgaact 2700
gaatgctgca tctgtatatt ttcctgctct gccttctggg actagccttg gcctaatcct 2760
tcctctagga gaagagcatt caggttttgg gagatggctc atagccaagc ccctctctct 2820
tagtgtgatc ccttggagca ccttcatgcc tggggtttct ctcccaaaag cttcttgcag 2880
tctaagcctt atcccttatg ttccccatta aaggaatttc aaaagacatg gagaaagttg 2940
ggaaggtttg tgctgactgc tgggagcaga atagccgtgg gaggcccacc aagcccttaa 3000
attcccattg tcaactcaga acacatttgg gcccatatgc caccctggaa caccagctga 3060
caccatgggc gtccacacct gctgctccag acaagcacaa agcaatcttt cagccttgaa 3120
atgtattatc tgaaaggcta cctgaagccc aggcccgaat atggggactt agtcgattac 3180
ctggaaaaag aaaagaccca cactgtgtcc tgctgtgctt ttgggcagga aaatggaaga 3240
aagagtgggg tgggcacatt agaagtcacc caaatcctgc caggctgcct ggcatccctg 3300
gggcatgagc tgggcggaga atccaccccg caggatgttc agagggaccc actccttcat 3360
ttttcagagt caaaggaatc agaggctcac ccatggcagg cagtgaaaag agccaggagt 3420
cctgggttct agtccctgct ctgcccccaa ctggctgtat aacctttgaa aaatcatttt 3480
ctttgtctga gtctctggtt ctccgtcagc aacaggctgg cataaggtcc cctgcaggtt 3540
ccttctagct ggagcactca gagcttccct gactgctagc agcctctctg gccctcacag 3600
ggctgattgt tctccttctc cctggagctc tctctcctga aaatctccat cagagcaagg 3660
cagccagaga agcccctgag agggaatgat tgggaagtgt ccacttctcc aaccggctca 3720
tcaaacacac tcctttgtct atgaatggca catgtaaatg atgttatatt ttgtatcttt 3780
tatatcatat gcttcaccat tctgtaaagg gcctctgcat tgttgctccc atcaggggtc 3840
tcaagtggaa ataaaccctc gta 3863

<210> 13
<211> 1129
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 5291211CB1

<400> 13
cgcggccctg caggcggttg cgttccccgt cgttaccctc tttctcttcc cgacgcgtga 60
gttaggccgt aatgccttgg ctgctctcag cccccaagct ggttcccgct gtagcaaacg 120
tccgcggcct ctcaggatgt atgttgtgtt cacagcgaag gtactccctt cagcctgtcc 180
cagaaaggag gattccaaac cgatacttag gccagcccag ccccttaca cacccacacc 240
tcctcagacc aggggaggta actccaggac tatctcaggt ggaatatgca cttcgcagac 300
acaaactaat gtctctgatc cagaaggaag ctcaagggca gagtgggaca gaccagacag 360
tggttgtgct ctccaaccct acatactaca tgagcaacga tattccctat actttccacc 420
aagacaacaa tttcctgtac ctatgtggat tccaagagcc tgatagcatt cttgtccttc 480
agagcctccc tggcaaacaa ttaccatcac acaaagccat acttttgtg cctcggcgag 540
atcccagtcg agaactttgg gatggtccgc gatctggcac tgatggagca atagctctaa 600
ctggagtaga cgaagcctat acgctagaag aatttcaaca tcttctacca aaaatgaaag 660
gtaacaaatg ggagcagaag tcacattaca aaccagattg ggattaaaac ctttcttgcc 720
tgtttagaca agtccttaat tttgttatcg tagcaccacc atgaaaagac catgagcacc 780
atgaaagaaa tgtaaagtct ttttcagaag attgagcttc tttaaagatt tcatttatgc 840
ctggtgcatt ggctcgcgcc cgtggtccca gcactttggg aggccggggc ggacaggttc 900
cttgtggccg ggagttcggg accagcctgg ccagcatggc ggagccccgt ctctgctaaa 960
aatagaaaaa ttagccaggc atggtcgcag gcgcctgtag tcccagctgc tcgggtggct 1020
caggcgggag aatcgcttga gcccgggagg tggaggttgc agtgatccga gatcatgcca 1080
ctgcattcca gcctaggcaa caaagcgaga ctgtcttaaa aaaaaaaaa 1129

<210> 14
<211> 2196
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature

<223> Incyte ID No: 7472661CB1

<400> 14
atgaggcagg cagaggcgcg ggtcaccctt agggcccccc tcttgctgct ggggctctgg 60
gtgctcctga ctccagtccg gtgttctcaa ggccatccct cgtggcacta cgcatcctcc 120
aaggtggtga ttcccaggaa ggagacgcac cacggcaaag accttcagtt tctgggctgg 180
ctgtcctaca gcctgcattt tgggggtcaa agacacatca ttcacatgcg gaggaaacac 240
cttctttggc ccagacatct gctggtgaca actcaggatg accaaggagc cttgcagatg 300
gatgacccct acatccctcc agactgctac tatctcagct acctggagga ggttcctctg 360
tccatggtca ccgtggacat gtgctgtggg ggcctcagag gcatcatgaa gctggacgac 420
cttgcctatg aaatcaaacc cctccaggat tcccgcaggc ttgaacatgt ttctcagata 480
gtggccgagc ccaacgcaac ggggcccaca tttagagatg gtgacaatga ggagacaaac 540
cccctgttct ctgaagcaaa tgacagcatg aatcccagga tatctaattg gctgtatagt 600
tctcatagag gcaatataaa aggctacgtt caatgttcca attcatattg tcgtgtagat 660
gacaatatta caacttgttc caaggaggtg gtccagatgt tcagtctcag tgacagcatt 720
gttcaaaata ttgatctgcg gtactatatt tatcttttga ccatatataa taattgtgac 780
ccagcccctg tgaatgacta tcgagttcag agtgcaatgt ttacctattt tagaacaacc 840
ttttttgata cttttcgtgt tcattcaccc acactactta ttaaagaggc accacatgaa 900
tgtaactatg aaccacaaag gtatagcttc tgtacacatt taggcctatt acacattggt 960
actctaggca gacattattt attagtagcc gtcataacaa cccagacact gatgagaagt 1020
actggtgaga agtacgatga taactactgc acatgtcaga aaagggcctt ctgcattatg 1080
cagcaatatc ctgggatgac agatgcgttc agtaactgtt cttatggaca tgcacaaaat 1140
tgttttgtac attcagcccg gtgtgttttc gaaacacttg ctcctgtgta taatgaaacc 1200
atgacaatgg ttcgctgtgg aaacctcata gcggatggga gggaggaatg tgactgtggc 1260
tccttcaagc agtgttatgc cagttattgc tgccgaagtg actgtcgctt aacaccgggg 1320
agcatctgtc atataggaga gtgctgtaca aactgcagct actccccacc agggactctc 1380
tgcagaccta tccaaaatat atgtgacctt ccagagtact gtcacgggac caccgtgaca 1440
tgccccgcaa acttttatat gcaagatgga accccgtgca ctgaagaagg ctactgctat 1500
catgggaact gcactgaccg caatgtgctc tgcaaggtaa tctttggtgt cagtgctgag 1560
gaggctcctg aggtctgcta tgacataaat cttgaaagtt accgatttgg acattgtact 1620
cgacgacaaa cagctctcaa caaccaggct tgtgcaggaa tagataagtt ttgtggaaga 1680
ctgcagtgta ccagtgtgac ccatcttccc cggctgcagg aacatgtttc attccatcac 1740
tcagtgacag gaggatttca gtgttttgga ctggatgacc accgtgcaac agacacaact 1800
gatgttgggt gtgtgataga tggcactcct tgtgttcatg gaaacttctg taataacacc 1860
aggtgcaatg cgactatcac ttcactgggc tacgactgtc gccctgagaa gtgcagtcat 1920
agaggggtgt gcaacaacag aaggaactgc cattgccata taggctggga tcctccactg 1980
tgcctaagaa gaggtgctgg tgggagtgtc gacagcgggc cacctccaaa aataacacgt 2040
tcggtcaaac aaagccaaca atcagtgatg tatctgagag tggtctttgg tcgtatttac 2100
accttcataa ttgcactgct ctttgggatg gccacaaatg tgcgaactat caggaccacc 2160
actgttaagg gatggacagt tactaaccct gaataa 2196

<210> 15
<211> 2837
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1795307CB1

<400> 15
ctgcggggca ggccgggggc agctgtctgt ctggctcttt ttgacagccc ccagtgcgaa 60
aggctgccag catgtcatca gtgagcccca tccagatccc cagtcgcctc ccgctgctgc 120
tcacccacga gggcgtcctg ctgcccggct ccaccatgcg caccagcgtg gactcggccc 180
gcaacctgca gctggtgcgg agccgccttc tgaagggcac gtcgctgcaa agcaccatcc 240
tgggcgtcat ccccaacacg cctgaccccg ccagcgacgc gcaggacctg ccgccgctgc 300
acaggattgg cacagctgca ctggccgttc aggttgtggg cagtaactgg cccaagcccc 360
actacactct gttgattaca ggcctatgcc gtttccagat tgtacaggtc ttaaaagaga 420
agccatatcc cattgctgaa gtggagcagt tggaccgact tgaggagttt cccaacacct 480
gtaaaatgag ggaggagcta ggagaactat cagagcagtt ttacaaatat gcagtacaat 540
tggttgaaat gttggatatg tctgtccctg cagttgctaa attgagacgt cttttagata 600
gtcttccaag ggaagcttta ccagacatct tgacatcaat tatccgaaca agcaacaaag 660
agaaactcca gattttagat gctgtgagcc tagaggagcg gttcaagatg actataccac 720
tgcttgtcag acaaattgaa ggcctgaaat tgcttcaaaa aaccagaaaa cccaagcaag 780
atgatgataa gagggttata gcaatacgcc ctattaggag aattacacat atctcaggta 840
ctttagaaga tgaagatgaa gatgaagata atgatgacat tgtcatgcta gagaaaaaaa 900

tacgaacatc tagtatgcca gagcaggccc ataaagtctg tgtcaaagag ataaagagac 960
tcaaaaaaat gcctcagtca atgccagaat atgctctgac tagaaattat ttggaactta 1020
tggtagaact tccttggaac aaaagtacaa ctgaccgcct ggacattagg gcagcccgga 1080
ttcttctgga taatgaccat tacgccatgg aaaaattgaa gaaaagagta ctggaatact 1140
tggctgtcag acagctcaaa aataacctga agggcccaat cctatgcttt gttggccctc 1200
ctggagttgg taaaacaagt gtgggaagat cagtggccaa gactctaggt cgagagttcc 1260
acaggattgc acttggagga gtatgtgatc agtctgacat tcgaggacac aggcgcacct 1320
atgttggcag catgcctggt cgcatcatca acggcttgaa gactgtggga gtgaacaacc 1380
cagtgttcct attagatgag gttgacaaac tgggaaaaag tctacagggt gatccagcag 1440
cagctctgct tgaggtgttg gatcctgaac aaaaccataa cttcacagat cattatctaa 1500
atgtggcctt tgacctttct caagttcttt ttatagctac tgccaacacc actgctacca 1560
ttccagctgc cttgttggac agaatggaga tcattcaggt tccaggttat acacaggagg 1620
agaagataga gattgcccat aggcacttga tccccaagca gctggaacaa catgggctga 1680
ctccacagca gattcagata ccccaggtca ccactcttga catcatcacc aggtatacca 1740
gagaggcagg ggttcgttct ctggatagaa aacttggggc catttgccga gctgtggccg 1800
tgaaggtggc agaaggacag cataaggaag ccaagttgga ccgttctgat gtgactgaga 1860
gagaaggttg cagagaacac atcttagaag atgaaaaacc tgaatctatc agtgacacta 1920
ctgacttggc tctaccacct gaaatgccga ttttgattga tttccatgct ctgaaagaca 1980
tccttgggcc cccgatgtat gaaatggagg tatctcagcg tttgagtcag ccaggagtag 2040
caataggttt ggcttggact cccttaggtg gagaaatcat gttcgtggag gcgagtcgaa 2100
tggatggcga gggccagtta actctgaccg gccagctcgg ggacgtgatg aaggagtccg 2160
cccacctcgc tatcagctgg ctccgcagca acgcaaagaa gtaccagctg accaatgctt 2220
ttggaagttt tgatcttctt gacaacacag acatccatct gcacttccca gctggagctg 2280
tcacaaaaga tggaccatct gctggagtta ccatagtaac ctgtctcgcc tcacttttta 2340
gtgggcggct ggtacgttca gatgtagcca tgactggaga aattacactg agaggtcttg 2400
ttcttccagt gggtggaatt aaagacaaag tgctggcggc acacagagcg ggactgaagc 2460
aagtcattat tcctcggaga aatgaaaaag accttgaggg aatcccaggc aacgtacgac 2520
aggatttaag ttttgtcaca gcaagctgcc tggatgaggt tcttaatgca gcttttgatg 2580
gtggctttac tgtcaagacc agacctggtc tgttaaatag caaactgtag gtccaaatct 2640
caatttttta gaattttaag ttatgaagtg ctcaaaggta ctgacacagt tgattttatt 2700
cacaccatta ggggtatgca agatgtccct gttttataaa cataatcaca acagtaataa 2760
acctcaagta gtggctagtg tttagtatag aaatataaga tgttgattta gtaaactgat 2820
aaaaatcgaa aaaaaaa 2837

<210> 16
<211> 716
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472685CB1

<400> 16
gtcgacttcg gcggggccca ggtgagaaag gcccacctgt gtcctggttg agggtctcca 60
gggttctttg gggcccgagg ccaatggtgg cagagtctac atagaactat gcttcgtggt 120
gttctgggga aaacctttcg acttgttggc tatactattc aatatggctg tatagctcat 180
tgtgcttttg aatacgttgg tggtgttgtc atgtgttctg gaccatcaat ggagcctaca 240
attcaaaatt cagatattgt ctttgcagaa aatcttagtc gacattttta tggtatccaa 300
agaggtgaca ttgtgattgc aaaaagccca agtgatccaa aatcaaatat ttgtaaaaga 360
gtaattggtt tggaaggaga caaaatcctc accactagtc catcagattt ctttaaaagc 420
catagttatg tgccaatggg tcatgtttgg ttagaaggtg acaatctaca gaattctaca 480
gattccaggt gctatggacc tattccatat ggactaataa gaggacgaat cttctttaag 540
atttggcctc tgagtgattt tggattttta cgtgccagcc ctaatggcca cagattttct 600
gatgattagt aagcatttat tcttttgact tgattattgt ctccttttca tgtgaattta 660
ttactcccgt tgaaaccgtg tacttaccaa taaactattt gctattcaaa aaaaaa 716

<210> 17
<211> 2894
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2777676CB1

<400> 17
cagaaggaat ggtagcagct cctctttgta catctggtag aattcggctg tgaatccatc 60
tggtcctgga ctttttttgg ttgaggtggc aggccactgc ccaatgctga agtcactgtt 120
taacgaaggt ggagcctgcc tttacctggt acacccatat aaggaaaagc ctgaggtcag 180
gagtaagcag acaccagcac tgctcttttct ccaagacggc cggccatgct ctcctcctct 240
gccagtctcc tccaccactc tctaacctga gagcctgtgg aacctgcccg tctcccctcc 300
tccatcagac acacctgcct aggaaacagg aaaggacctc ggaagtcttc taaggagagt 360
catggcgtat taccaggagc cttcagtgga gacctccatc atcaagttca aagaccagga 420
ctttaccacc ttgcgggatc actgcctgag catgggccgg acgtttaagg atgagacatt 480
ccctgcagca gattcttcca taggccagaa gctgctccag gaaaaacgcc tctccaatgt 540
gatatggaag cggccacagg atctaccagg gggtcctcct cacttcatcc tggatgatat 600
aagcagattt gacatccaac aaggaggcgc agctgactgc tggttcctgg cagcactggg 660
atccttgact cagaacccac agtacaggca gaagatcctg atggtccaaa gcttttcaca 720
ccagtatgct ggcatttttcc gtttccggtt ctggcaatgt ggccagtggg tggaagtggt 780
gattgatgac cgcctacctg tccagggaga taaatgcctc tttgtgcgtc ctcgccacca 840
aaaccaagag ttctggccct gcctgctgga gaaggcctat gccaagctgc tcggatccta 900
ttccgatctg cactatggct tcctcgagga tgccctggtg gacctcacag gaggcgtgat 960
caccaacatc catctgcact cttcccctgt ggacctggtg aaggcagtga agacagcgac 1020
caaggcaggc tccctgataa cctgtgccac tccaagtggg ccaacagata cagcacaggc 1080
gatggagaat gggctggtga gtctccatgc ctacactgtg actggggctg agcagattca 1140
ataccgaagg ggctgggaag aaattatctc cctgtggaac ccctggggct ggggcgagac 1200
cgaatggaga gggcgctgga gtgatgggtc tcaggagtgg gaggaaacct gtgatccgcg 1260
gaaaagccag ctacataaga aacgggaaga tggcgagttt tggatgtcat gtcaagattt 1320
ccaacagaaa ttcatcgcca tgtttatatg tagcgaaatt ccaattaccc tggaccatgg 1380
aaacacactc cacgaaggat ggtcccaaat aatgtttagg aagcaagtga ttctaggaaa 1440
cactgcagga ggacctcgga atgatgctca attcaacttc tctgtgcaag agccaatgga 1500
aggcaccaat gttgtcgtgt gcgtcacagt tgctgtcaca ccatcaaatt tgaaagcaga 1560
agatgcaaaa tttccactcg atttccaagt gattctggct ggctcacagc ggttccggga 1620
gaaatttcca cccgtgtttt tttcctcgtt cagaaacact gtccaaagct caaataataa 1680
attccgccgc aacttcacca tgacttacca tctgagccct gggaactatg ttgtggttgc 1740
acagacacgg agaaaatcag cggagttctt gctccgaatc ttcctgaaaa tgccagacag 1800
tgacaggcac ctgagcagcc atttcaacct cagaatgaag ggaagccctt cagaacatgg 1860
ctcccaacaa agcattttca acagatatgc tcagcagagg ctggacattg atgccaccca 1920
gcttcagggc cttctcaacc aggagcttct aacaggacct ccaggggaca tgttctcctt 1980
agatgagtgc cgcagcttgg tggctctgat ggaactgaaa gtgaatgggc ggctagacca 2040
agaggagttt gcgcgactgt ggaagcgcct tgttcactac cagcatgttt tccagaaggt 2100
tcagacaagc cctggagtcc tcctgagctc ggacttgtgg aaggccatag agaatacaga 2160
cttcctcaga gggatcttca tcagccgtga gctgctgcat ctggtgaccc tcaggtacag 2220
cgacagcgtc ggcagggtca gcttccccag cctggtctgc ttcctgatgc ggcttgaagc 2280
catggcaaag accttccgca acctctctaa ggatggaaaa ggactctacc tgacagaaat 2340
ggagtggatg agcctggtca tgtacaactg aagcaaagag gaaagcagac ccatggctca 2400
ggacaagctc ccagtgatca ctcaagaatc tggctctcat tctaagaggc tgtgctgccc 2460
agtatggtgg ttgtgataaa tctaaaccag ccctgcatga aacagagtcc aagctgtctc 2520
ccaacagcct gggttcggtc cttggctggc ccaggcccag ttaagcctgt ggccaccaag 2580
cagctcatct gagcactttg ggatgtattc agcctacgtt gccctggaaa aggaagcagg 2640
agatgtctcc ctgtgggaaa ggagaagaga agttgtctct gagtcccctg tcaccagttg 2700
gattcatttc ttggaagagc cagaatgagc cactttgacc accctcgggt gctatgggtg 2760
acacaagagc tgtccactgg gtgtttgcag aataattaca ctatcttatg tctggatcct 2820
gatgatttca cagctaaatg gcaaaaataa aacatgtttc ccataaaaaa aaaaaaaaaa 2880
aaaaaaaaaa aaaa 2894

<210> 18
<211> 1444
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473620CB1

<400> 18
atgcagggca cccctggagg cgggacgcgc cctgggccat cccccgtgga caggcggaca 60
ctcctggtct tcagctttat cctggcagca gctttgggcc aaatgaattt cacaggggac 120
caggttcttc gagtcctggc caaagatgag aagcagcttt cacttctcgg ggatctggag 180
ggcctgaaac cccagaaggt ggacttctgg cgtggcccag ccaggcccag cctccctgtg 240
gatatgagag ttcctttctc tgaactgaaa gacatcaaag cttatctgga gtctcatgga 300

cttgcttaca gcatcatgat aaaggacatc caggtgctgc tggatgagga aagacaggcc 360
atggcgaaat cccgccggct ggagcgcagc accaacagct tcagttactc atcataccac 420
accctggagg agatatatag ctggattgac aactttgtaa tggagcattc cgatattgtc 480
tcaaaaattc agattggcaa cagctttgaa aaccagtcca ttcttgtcct gaagttcagc 540
actggaggtt ctcggcaccc agccatctgg attgacactg gaattcactc ccgggagtgg 600
atcacccatg ccaccggcat ctggactgcc aataagattg tcagtgatta tggcaaagac 660
cgtgtcctga cagacatact gaatgccatg gacatcttca tagagctcgt cacaaaccct 720
gatgggtttg cttttaccca cagcatgaac cgcttatggc ggaagaacaa gtccatcaga 780
cctggaatct tctgcatcgg cgtggatctc aacaggaact ggaagtcggg ttttggagga 840
aatggttcta acagcaaccc ctgctcagaa acttatcacg ggccctcccc tcagtcggag 900
ccggaggtgg ctgccatagt gaacttcatc acagcccatg gcaacttcaa ggctctgatc 960
tccatccaca gctactctca gatgcttatg tacccttacg gccgattgct ggagcccgtt 1020
tcaaatcaga gggagttgta cgatcttgcc aaggatgcgg tggaggcctt gtataaggtc 1080
catgggatcg agtacatttt tggcagcatc agcaccaccc tctatgtggc cagtgggatc 1140
accgtcgact gggcctatga cagtggcatc aagtacgcct tcagctttga gctccgggac 1200
actgggcagt atggcttcct gctgccggcc acacagatca tccccacggc ccaggagacg 1260
tggatggcgc ttcggaccat catggagcac accctgaatc acccctacta gcagcacgac 1320
tgagggcagg aggctccatc cttctcccca aggtctgtgg ctcctcccga aacccaagtt 1380
atgcatcccc atccccatgc cctcatcccg acctttttaga aaataaatac aagtttgaac 1440
aggc 1444

<210> 19
<211> 1872
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 3405540CB1

<400> 19
ctgcgtcaga aggttctaac ttttgtcatc actattacca gcattgtcat cgttatcgtt 60
atcttcgtca tcatcattac caccgttata cctgatactg ccataacaat cagaacatta 120
tgtacaggca cggcatatct tcccaaagat cttggccact atggactacg atctttattt 180
ttcttggagt ggcggcaatc ttgggagtaa ccattggtct tcttgttcat tttctggcag 240
ttgagaagac ttactattat caaggtgatt ttcatatttc tggagtcaca tacaatgata 300
attgtgaaaa cgcagcttca caagccagca caaatctaag caaagatatt gagactaaga 360
tgttaaatgc atttcaaaat tccagtatat ataaggaata tgtcaaatct gaggtcatca 420
aacttctgcc taatgccaat ggttcaaatg tgcagttaca gctgaaattc aagtttcctc 480
cagcagaagg agttagcatg aggactaaaa tcaaggctaa attacatcag atgttgaaaa 540
acaacatggc atcctggaat gcagttcctg cttccattaa actcatggaa atcagcaagg 600
ctgcttctga aatgcttacc aacaactgtt gtgggagaca agtagccaac agtatcataa 660
ctggcaacaa aattgtgaat ggaaaaagct ccctggaggg ggcatggcca tggcaggcca 720
gcatgcaatg gaaaggccgt cactactgtg gagcctctct gatcagcagc aggtggctat 780
tatctgcagc tcactgcttt gctaagaaaa ataattcaaa agattggact gtcaactttg 840
gagttgtagt aaataaacca tatatgacac ggaaagtcca aaacattatt tttcatgaaa 900
attatagcag tcctgggctt catgatgata ttgcccttgt gcagcttgct gaagaagttt 960
cttttacaga gtacattcgt aagatttgtc ttcctgaagc caaaatgaag ctctcagaaa 1020
atgacaatgt tgtagttaca ggttggggaa cactttatat gaatggttca tttccagtga 1080
tacttcaaga agccttttttg aagattattg acaacaaaat ttgcaatgcc tcatatgcat 1140
actctggctt tgtgactgat tcaatgttat gtgctggatt tatgtcagga gaagctgatg 1200
catgtcagaa tgattctggt ggaccactag cttaccctga ttccagaaat atctggcatc 1260
ttgttggaat agtaagctgg ggtgatggat gtggtaaaaa gaataagcca ggtgtctata 1320
ctcgagtgac ttcttatcgc aattggatta catccaagac tggactctga aaaaaaagga 1380
attatacaaa ggaacataaa gaccactgta ggctatcctt atctgtggct ttgggatctt 1440
tttattttga ttgttggtaa actaatcatt ttattatata taaaaatata aatgtcttta 1500
ttaatatttt attatatcta aaacactgtc catgcattta ttaatctttt aaattaatgt 1560
aagagatagg aagaggccag gtgtagtggc tcatgcctgt aatcctagca ctttgagaga 1620
ccgagatagg cggattgctt gagcccagga gtttgagggc agcctgggaa acatggctaa 1680
atccgtgttc aaaaaataca aaaattagct aggcatggtg gcgtgcacct atagtcccag 1740
ctattcagga ggctgaggtg ggaggattgc ttgagcccag caggttgagg ctgtggtgat 1800
ttgtgattgc gccaccactg cactccagcc tgggcaacag agtgagaccc tgtctcaaaa 1860
aaaaaaaaaa aa 1872

<210> 20
<211> 2605

<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4069067CB1

<400> 20
ttctcaagca cttctgctct cctctaccag aatcactcag aatgcttccc gggtgtatat 60
tcttgatgat tttactcatt cctcaggtta aagaaaagtt catccttgga gtagagggtc 120
aacaactggt tcgtcctaaa aagcttcctc tgatacagaa gcgagatact ggacacaccc 180
atgatgatga catactgaaa acgtatgaag aagaattgtt gtatgaaata aaactaaata 240
gaaaaacctt agtccttcat cttctaagat ccagggagtt cctaggctca aattacagtg 300
aaacattcta ctccatgaaa ggagaagcgt tcaccaggca tcctcagatc atggaccact 360
gttactataa aggaaacatc ctaaatgaaa agaattctgt tgccagcatc agtacttgtg 420
acgggttgag gggattcttc agaataaacg accaaagata cctcattgaa ccagtgaaat 480
actcagatga gggagaacat ttggtgttca aatataacct gagggtgccg tatggtgcca 540
attattcctg tacagagctt aattttacca gaaaaactgt tccaggggat aatgaatctg 600
aagaagactc caaaataaaa ggcatccatg atgaaaagta tgttgaattg ttcattgttg 660
ctgatgatac tgtgtatcgc agaaatggtc atcctcacaa taaactaagg aaccgaattt 720
ggggaatggt caattttgtc aacatgattt ataaaacctt aaacatccat gtgacgttgg 780
ttggcattga aatatggaca catgaagata aaatagaact atattcaaat atagaaacta 840
ccttattgcg tttttcattt tggcaagaaa agatccttaa aacacggaag gattttgatc 900
atgttgtatt actcagtggg aagtggctct actcacatgt gcaaggaatt tcttatccag 960
ggggtatgtg cctgccctat tattccacca gtatcattaa ggatctttta cctgacacaa 1020
acataattgc aaacagaatg gcacatcaac tggggcataa ccttgggatg cagcatgacg 1080
agttcccatg cacctgtcct tcaggaaaat gcgtgatgga cagtgatgga agcattcctg 1140
cactggacct cagtaaatgc agacaaaacc aataccacca gtacttgaag gattataagc 1200
caacatgcat gctcaacatt ccatttcctt acaatttca tgatttccaa ttttgtggaa 1260
acaagaagtt ggatgagggt gaagagtgtg actgtggccc tgctcaggag tgtactaatc 1320
cttgctgtga tgcacacaca tgtgtactga agccaggatt tacttgtgca gaaggagaat 1380
gctgtgaatc ttgtcagata aaaaaagcag ggtccatatg cagaccggcg aaagatgaat 1440
gtgattttcc tgagatgtgc actggccact cgcctgcctg tcctaaggac cagttcaggg 1500
tcaatggatt tccttgcaag aactcagaag gctactgttt catggggaaa tgtccaactc 1560
gtgaggatca gtgctctgaa ctatttgatg atgaggcaat agagagtcat gatatctgct 1620
acaagatgaa tacaaaagga aataaatttg gatactgcaa aaacaaggaa aacagatttc 1680
ttccctgtga ggagaaagat gtcagatgtg gaaagatcta ctgcactgga ggggagcttt 1740
cctctctcct tggagaagac aagacttatc accttaagga tccccagaag aatgctactg 1800
tcaaatgcaa aactattttt ttataccatg attctacaga cattggcctg gtggcgtcag 1860
gaacaaaatg tggagaggga atggtgtgca acaatggtga atgtctaaac atggaaaagg 1920
tctatatctc aaccaattgc ccctctcagt gcaatgaaaa tcctgtggat ggccacggac 1980
tccagtgcca ctgtgaggaa ggacaggcac ctgtagcctg tgaagaaacc ttacatgtta 2040
ccaatatcac catcttggtt gttgtgcttg tcctggttat tgtcggtatc ggagttctta 2100
tactattagt tcgttaccga aaatgtatca agttgaagca agttcagagc ccacctacag 2160
aaaccctggg agtggagaac aaaggatact ttggtgatga gcagcagata aggactgagc 2220
caatcctgcc agaaattcat ttcctaaatc agagaactcc agaatccttg gaaagcctgc 2280
ccactagttt ttcaagtccc cactacatca cactgaaacc tgcaagtaaa gattcaagag 2340
gaatcgcaga tcccaatcaa agtgccaagt gagcttgaag ttggatatcc aaaatggccg 2400
tgcaagctta ggctggggat tctggatgca acgtctttac aaccttacct agatatctgc 2460
tactcacatt tttggtagtg tttcaaacgt tctttatcca gacagacaat gtttaagaga 2520
aacaacttat ttctgttaat atttaccggt agaattcaca ccctctatca taaacatatg 2580
ctgcagaaaa aaaaaaaaaa aaaaa 2605

<210> 21
<211> 1896
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 4317947CB1

<400> 21
cacagagacg gggtttctca gtgttgtcca ggctggtctc gaactcctga gctcaagcag 60
tccgcccacc ttgacctccc aaagtgctgg gattacaggc tttcatagaa accatgttca 120
ccagtaaagc ccctgtggaa gaagcctttc tttatgctaa gtttgaattt gaatgccggg 180

ctcgtggcgc agacatttta gcctatccac ctgtggtggc tggtggtaat cggtcaaaca 240
ctttgcacta tgtgaaaaat aatcaactca tcaaggatgg ggaaatggtg cttctggatg 300
gaggttgtga gtcttcctgc tatgtgagtg acatcacacg tacgtggcca gtcaatggca 360
ggttcaccgc acctcaggca gaactctatg aagccgttct agagatccaa agagattgtt 420
tggccctctg cttccctggg acaagcttgg agaacatcta cagcatgatg ctgaccctga 480
taggacagaa gcttaaagac ttggggatca tgaagaacat taaggaaaat aatgccttca 540
aggctgctcg aaaatactgt cctcatcatg ttggccacta cctcgggatg gatgtccatg 600
acactccaga catgccccgt tccctccctc tgcagcctgg gatggtaatc acaattgagc 660
ccggcattta tattccagag gatgacaaag atgccccaga gaagtttcgg ggtcttggtg 720
tacgaattga ggatgatgta gtggtgactc aggactcacc tctcatcctt tctgcagact 780
gtcccaaaga gatgaatgac attgaacaga tatgcagcca ggcttcttga ccttcactgc 840
ggcccacatg cacctcaggt tcaaaatggg tgtcttctgg cagccctgca cgtgtgcttt 900
ctgagtgtct ctgtgtgtgc attaatatat gcattccatt tgggagcata gcagctgtgt 960
gaatgtatgt aattgtgtgt gggggggttt ttgttttaag tagttagaag tctgggaaaa 1020
tgaatttttg aatagtatgt tactgcagct ttggtaacat taattctata gaattaatga 1080
tcagagcaag tttaattttt aaacataaag gtcttggtta cacatgtcca tgcattccag 1140
ttaacacatt taaacatata gaaaattcac ttcttctttc aagtccttcc ctttctatac 1200
ttttgctgag atcaacccaa tatcataaga agtttgtgtg atgcctgtat ttttagctaa 1260
ttaccaaggt ctctgcctat gcaaaaattc atctttcggt gattgtgggc ggccaataca 1320
taccttctgt aactcctccc tacctattct agatcctgca tatctactgg agccatatga 1380
gccctcaggg gcactggtgt cctgcagtct tactagtcag ccacagtcca ggttccagca 1440
acactagcta cctaccaaac aaaacacttc tgagcttttc ttgccatttg gttgcctggg 1500
acctggaaag tgtttagatt cttgataaag atggaaatgg aagagaaaga aaattatata 1560
tgtatatact aaagatccag acatcagggc tgggcacagc ggttccgcct gtagtcccag 1620
cactttggga ggtccaggcg ggtggatctc ttgagaccag gaggtcgaga ccagcctggc 1680
cggcgtggtg gagcactgtc tctactaaaa atgcaaaaat gagctgggtg tgatgataca 1740
tgcctgtggt cctggctact tgggaggctg aggcatgaga attgcttgaa cctgggaggt 1800
ggaggttgca gtgagctgag ctcatgcgac tgcactccag cctgggtgac tgagcgagac 1860
tctgcctcat aaatacaaat aaataaatac acctcc 1896

<210> 22
<211> 997
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473303CB1

<400> 22
atttaatatg atcactatag ggaatttggc cctcgaggcc aagaattcgg cacgaggcaa 60
aaaggagacc agacaggagg cgtctgtaga gatatcatga acttcaactt agctttgttt 120
tccagagact ggagctaaac tgggctttca acatcatcat gaagtttatc ctcctctggg 180
ccctcttgaa tctgactgtt gctttggcct ttaatccaga ttacacagtc agctccactc 240
ccccttactt ggtctatttg aaatctgact acttgccctg cgctggagtc ctgatccacc 300
cgctttgggt gatcacagct gcacactgca atttaccaaa gcttcgggtg atattggggg 360
ttacaatccc agcagactct aatgaaaagc atctgcaagt gattggctat gagaagatga 420
ttcatcatcc acacttctca gtcacttcta ttgatcatga catcatgcta atcaagctga 480
aaacagaggc tgaactcaat gactatgtga aattagccaa cctgccctac caaactatct 540
ctgaaaatac catgtgctct gtctctacct ggagctacaa tgtgtgtgat atcggaagtt 600
tgacttctat cttttcttta gacaaagagc ccgattcact gcaaactgtg aacatctctg 660
taatctccaa gcctcagtgt cgcgatgcct ataaaaccta caacatcacg gaaaatatgc 720
tgtgtgtggg cattgtgcca ggaaggaggc agccctgcaa ggaagtttct gctgccccgg 780
caatctgcaa tgggatgctt caaggaatcc tgtcttttgc ggatggatgt gttttgagag 840
ccgatgttgg catctatgcc aaaatttttt actatatacc ctggattgaa aatgtaatcc 900
aaaataactg agctgtggca gttgtggacc atatgacaca gcttgtcccc atcgttcacc 960
tttagaatta aatataaatt aactcctcaa aaaaaaa 997